
Defining the three-dimensional mechanisms of context-specific co-regulation in *Drosophila*

Joseph Louis Aguilera

BS, Bioengineering, University of California, Santa Cruz, 2020

MS, Data Science, Brown University, 2024

Department of Molecular Biology, Cell Biology, and Biochemistry

Data Science Institute

Brown University

Providence, Rhode Island

May 2026

© May 2026 Joseph Louis Aguilera
All Rights Reserved

This dissertation by **Joseph Louis Aguilera** 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 _____

Dr. Erica Larschan, Advisor

Recommended to the Graduate Council

Date _____

Dr. Kate O'Connor-Giles, Committee Chair

Date _____

Dr. Nicola Neretti, Reader

Date _____

Dr. Nicolas Fawzi, Reader

Date _____

Dr. Melissa Harrison, External Reader

Approved by the Graduate School

Date _____

Dr. David P. Lindstrom, Dean of the Graduate School

Joseph Louis Aguilera

joseph_aguilera@brown.edu | www.linkedin.com/in/josephlouis2025

EDUCATION

Brown University

PhD, Molecular Biology, Cell Biology, and Biochemistry
MS, Data Science

2026
2024

University of California, Santa Cruz

BS, Bioengineering (Biomolecular Concentration) with Bioinformatics Minor

2020

RESEARCH EXPERIENCE

Doctoral Thesis Research

May 2021 – Present

Advisor: Erica Larschan, PhD

Topic: Using Multi-omics and Machine Learning to Reveal Chromatin Domain Formation in *Drosophila*

Research: My research centers on developing and applying machine learning models to high-dimensional biological systems, with an emphasis on learning predictive and interpretable structure from noisy, multi-modal data. I integrate deep learning, graph-based modeling, and representation learning with large-scale functional genomics and 3D chromatin data to understand how regulatory signals propagate through complex networks. I design interpretable architectures, including graph neural networks and sequence-based pre-trained models, to predict gene expression and chromatin interactions from sequence, epigenomic, and spatial features. These models operate on millions of observations and heterogeneous inputs, requiring careful feature engineering, regularization, and evaluation under class imbalance. As a primary application domain, I study how transcription factors shape genome organization in *Drosophila*, using this system as a controlled testbed for modeling causality, competition, and network dynamics. I analyze high-resolution chromatin conformation (Micro-C, Hi-ChIP) and RNA data (SLAM-seq) to quantify how signals compete and dominate within a network. Overall, my work emphasizes model interpretability, generalization, and robustness, translating complex biological signals into structured, explainable predictions.

PUBLICATIONS

Aguilera, J.L.*, Cortez, K.*, Segarra Alonzo, L.C., Aldana, M., Aragon, A., Gray, C., Woodman-Sousa, M., Grive, K., Ray, M., & Larschan, E. (2026). Differential chromatin looping regulated by two GA-binding transcription factors creates an X-specific chromatin environment for dosage compensation. *In Review (Genome Research)*. *Authors contributed equally

Aguilera, J.L., Duan, E., Cortez, K., Lee, R., Aragon, A., Ray, M., & Larschan, E. (2026). The CLAMP GA-binding transcription factor regulates heat stress-induced transcriptional repression. *eLife*.

Aguilera, J.L.*, Mahableshwarkar, P.*, Zhang, C., Tiffany, C., & Larschan, E. (2026). Bind-GPS: Predicting and interpreting CLAMP binding sites using multi-omic data and 3D chromatin interactions. *In Preparation*. *Authors contributed equally

Ray, M., Conard, A. M., Urban, J., Mahableshwarkar, P., **Aguilera, J. L.**, Huang, A., Vaidyanathan, S., & Larschan, E. (2023). Sex-specific splicing occurs genome-wide during early *Drosophila* embryogenesis. *eLife*.

Johnson, L. C., Vo, A. A., Clancy, J. C., Myles, K. M., Pooranachithra, M., **Aguilera, J. L.**, Levenson, M. T., Wohlenberg, C., Rechtsteiner, A., Ragle, J. M., Chisholm, A. D., & Ward, J. D. (2023). NHR-23 activity is necessary for *C. elegans* developmental progression and apical extracellular matrix structure and function. *Development*.

Kumar, A. V., Mills, J., Parker, W. M., Leitão, J. A., Rodriguez, D. I., Daigle, S. E., Ng, C., Patel, R., **Aguilera, J. L.**, Johnson, J. R., Wong, S. Q., & Lapierre, L. R. (2023). Lipid droplets modulate proteostasis, SQST-1/SQSTM1 dynamics, and lifespan in *C. elegans*. *iScience*.

Martinez, M. A. Q., Kinney, B. A., Medwig-Kinney, T. N., Ashley, G., Ragle, J. M., Johnson, L., **Aguilera, J. L.**, Hammell, C. M., Ward, J. D., & Matus, D. Q. (2020). Rapid Degradation of *Caenorhabditis elegans* Proteins at Single-Cell Resolution with a Synthetic Auxin. *G3 Genes|Genomes|Genetics*.

FELLOWSHIPS & GRANTS

NIH ARC F99/K00 - Predoctoral to Postdoctoral Fellow Transition Award (Discontinued)

January 2025

[Howard Hughes Medical Institute Gilliam's Fellowship](#)

September 2022 – September 2025

The program provides awards to pairs of dissertation advisers and their graduate students based on what HHMI values and considers essential components of the environment, particularly the institution and adviser's commitment to creating a healthy academic ecosystem and the student's potential for scientific leadership.

Brown University Open Graduate Education Fellowship

September 2022 – June 2024

The Open Graduate Education Program allows select Brown doctoral students to pursue a master's degree in a secondary field. The program is designed to enable select doctoral students to write unique dissertations, pursue research, and forge career pathways that require knowledge in more than one field.

Brown University Diversity Fellowship

September 2020 – June 2023

The Diversity Fellowship of the University and Graduate School Diversity and Inclusion Action Plans (DIAPs) focuses on diversifying doctoral programs across all disciplines. The fellowship provides three years of support.

Brown University MCB NIH T32 training grant Awardee

September 2020 – September 2022

NIH–Maximizing Access to Research Careers (MARC) Fellowship

June 2018 – June 2020

The MARC fellowship aspires to increase the nation's pool of students from underrepresented groups who have the research experience and science preparation to succeed in biomedical Ph.D. programs. MARC Fellows are supported through their academic and research education with two years of financial support and a summer research experience at a research-intensive institution.

DATA SCIENCE MASTER'S PROJECTS

Bind-GPS: Predicting and interpreting CLAMP binding sites using multi-omic data and 3D chromatin interactions

December 2023

- BindGPS (SeqGPS & ChromGPS) predicts and interprets CLAMP transcription factor (TF) binding sites in *Drosophila* using DNA sequences, histone modifications, and 3D chromatin interactions.
- SeqGPS (DNABERT model) predicts MSL Recognition Elements (MREs) from DNA.
- ChromGPS (GCN model) integrates epigenomic & 3D chromatin data to classify MREs.
- This project grants insight into X-chromosome dosage compensation with potential applications in cancer research and therapeutic design.

Targeting the *Drosophila* X-chromosome using Transformers

June 2023

- Extracted publicly available *Drosophila* ChIP-seq data from the Gene Expression Omnibus.
- Using transcription factor binding sites (from ChIP-seq), I used a transformer deep learning architecture to reveal DNA patterns which exhibited high specificity to the *Drosophila* X-chromosome.
- Understanding how DNA sequences drive transcription factor binding could provide novel insight into the design of future therapeutics.

Using Supervised Machine Learning to Diagnose Breast Cancer from Fine Needle Aspiration Data

December 2022

- Using the Breast Cancer Wisconsin Dataset, I performed an analysis on an array of machine learning algorithms to determine which was best for diagnosing breast cancer.
- The analysis underscored the power of classic machine learning methods in diagnostic practices.

ORAL & POSTER PRESENTATIONS

Oral Presentation (Keynote)

March 7, 2026

2026 SACNAS UPR-RP 3rd Local Conference

Aguilera, J.L.*, Cortez, K.*, Segarra Alonzo, L.C., Aldana, M., Aragon, A., Gray, C., Woodman-Sousa, M., Grive, K., Ray, M., & Larschan, E. Differential chromatin looping regulated by two GA-binding transcription factors creates an X-specific chromatin environment for dosage compensation
San Juan, PR.

Oral Presentation

October 30, 2025

2025 SACNAS National Diversity in STEM Conference

Aguilera, J.L., Duan, E., Cortez, K., Lee, R., Aragon, A., Ray, M., Larschan, E. The CLAMP GA-binding transcription factor regulates heat stress-induced transcription repression by associating with 3D chromatin loops
Columbus, OH.

Poster Presentation

June 15, 2025

2025 Genome Architecture in Cell Fate and Disease

Aguilera, J.L., Duan, E., Cortez, K., Lee, R., Aragon, A., Ray, M., Larschan, E. The CLAMP GA-binding transcription factor regulates heat stress-induced transcription repression by associating with 3D chromatin loops
Ventura, CA.

Oral Presentation

June 6, 2024

2024 Nuclear Bodies: Hubs of Genomic Activity

Aguilera J. L., Ray M., Larschan E. Clamp-Mediated Chromatin Domain Formation Affects Transcription.
Niagara Falls, NY.

Poster Presentation

March 13, 2023

2023 Epigenetics, Chromatin, Development and Disease

Aguilera J. L., Ray M., Larschan E. X marks the spot: Targeting an active chromatin domain to the X chromosome.
Victoria, British Columbia.

Poster Presentation

November 10, 2022

2022 CSHL Biological Data Science Meeting

Aguilera J. L., Ray M., Larschan E. X marks the spot: Targeting an active chromatin domain to the X chromosome.
Cold Spring Harbor, NY.

Oral Presentation

June 18, 2022

2022 New England Regional SACNAS Meeting

Aguilera J. L., Ray M., Larschan E. X marks the spot: Targeting an active chromatin domain to the X chromosome.
Providence, RI.

Oral Presentation

April 7, 2022

63rd Annual *Drosophila* Research Conference

Aguilera J. L., Ray M., Larschan E. X marks the spot: Targeting an active chromatin domain to the X chromosome.
San Diego, CA.

Poster Presentation

April 9, 2022

63rd Annual *Drosophila* Research Conference

Aguilera J. L., Ray M., Larschan E. X marks the spot: Targeting an active chromatin domain to the X chromosome.
San Diego, CA.

Poster Presentation

November 1, 2019

2019 SACNAS National Diversity in STEM Conference

Aguilera J., Johnson L., Ward J.D. *Characterizing NHR-23 activity in the Caenorhabditis elegans molting cycle using fluorescence microscopy.*
Honolulu, HI.

Poster Presentation

August 16, 2019

UC Santa Cruz 10th Annual Physical and Biological Summer Research Symposium

Aguilera J., Johnson L., Ward J.D. *Characterizing NHR-23 activity in the Caenorhabditis elegans molting cycle using fluorescence microscopy.*
Santa Cruz, CA.

Poster Presentation

November 15, 2018

2018 Annual Biomedical Research Conference for Minority Students

Aguilera J., Johnson L., Ward J.D. *Determining the timing and localization of NHR-23 and NHR-85 in the Caenorhabditis elegans molt using fluorescence microscopy.*
Indianapolis, IN.

Poster Presentation

August 17, 2018

UC Santa Cruz 9th Annual Physical and Biological Summer Research Symposium

Aguilera J., Johnson L., Ward J.D. *Determining the timing and localization of NHR-23 and NHR-85 in the Caenorhabditis elegans molt using fluorescence microscopy.*
Santa Cruz, CA.

PROFESSIONAL LEADERSHIP & MEMBERSHIP

Brown University SACNAS Chapter: President

September 2023 – May 2025

Brown University SACNAS Chapter: Undergraduate Research Chair

September 2022 – September 2023

Curriculum vitae

Brown University SACNAS Chapter: Secretary	September 2020 – September 2022
2022 New England Regional SACNAS Meeting Organizing Committee	June 2021 – June 2022
UC Santa Cruz SACNAS Chapter: Active Member	January 2019 – June 2020
National Society of Collegiate Scholars Honors Society (NSCS)	September 2017 – June 2020

SKILLS

- Python (Advanced), UNIX (Advanced), High-performance computing (Advanced), R (Proficient), Java (Basic)
- PyTorch (Advanced), TensorFlow (Proficient)
- Fluorescence Microscopy (Advanced), Light Microscopy (Advanced)
- BIOVIA's Discovery Studio (Proficient)
- Graphing Software: Gnuplot (Advanced), MATLAB (Proficient), Matplotlib (Advanced), Seaborn (Advanced)
- Languages: English (Fluent), Spanish (Basic)

RELEVANT COURSES

- Deep Learning
- Deep Learning in Genomics
- Machine Learning: from Theory to Algorithms
- Hands-on Machine Learning
- Statistical Learning
- Data Engineering
- Advanced Introduction to Probability for Computing and Data Science
- Data and Society

ACTIVITIES & INVOLVEMENT

Assistant Junior Varsity Football Coach **June 2017 – October 2019**
 Coached Channel Islands High students specializing in the running-back and line-backer positions.

Acknowledgments

First, I would like to thank my PhD advisor, Dr. Erica Larschan, for her endless support. From science to mentorship, in every chat, I was always being inspired by her character and learning something new. To the members of my committee, Dr. O'Connor-Giles, Dr. Nicola Neretti, and Dr. Nicolas Fawzi, thank you for your guidance and support through my years in the graduate program. To Dr. Ritambhara Singh and Dr. Andras Zsom, thank you for building my foundation in machine learning. Through your courses, your excellent lectures granted me the intuition to continue learning and refining my skills in the field.

To my fellow lab-mates in the Larschan laboratory, thank you for all the laughs and the feedback, it was a pleasure getting to know and learn from you all. To my mentees, teaching y'all were the best times I had in graduate school.

To the HHMI Gilliam community, thank you for accepting me as part of the community. At every retreat, I was pleasantly reminded that the future of science is in great hands, as you all will become forces in your respective fields.

To my professors and mentors at the University of California, Santa Cruz, thank you for love and support when I commenced my pursuit in higher education. To Yulianna Ortega, thank you for seeing the scientist in me when I could not and inspiring me to pursue this path.

To my wonderful partner, Sydney Fregoso, thank you for your unwavering support through these difficult years. You witnessed every up and down and always had the right words to say that would ground me.

To my father, you would have been proud of what I have become. From our last conversation, I have been determined to become the very best I could be. Thank you for teaching me how to be clever. Thank you for teaching me how to confront my fears. Thank you for being the memory which grants me the greatest strength.

And finally, to my mother, words cannot express how grateful I am to be your son. Thank you for believing in me from the start. Thank you for teaching me resignation is not an option. Thank you for accepting no less than excellence, as you are solely responsible for building the drive I have today.

Table of Contents

Signature Page	iii
Curriculum Vitae	iv
Acknowledgments	viii
List of Figures	xiii
List of Tables	xiv
Abstract	xv
1 Chapter 1: Introduction	1
1.1 Co-regulation of the <i>Drosophila melanogaster</i> genome	2
1.1.1 What is transcriptional co-regulation and why does it matter?	2
1.1.2 <i>Drosophila melanogaster</i> as a model for studying co-regulation	2
1.1.3 Two paradigms of precise co-regulation in <i>Drosophila</i>	3
1.1.4 GA-binding pioneer transcription factors as central regulators	5
1.2 Chromatin organization and transcriptional regulation	6
1.2.1 Hierarchical organization of eukaryotic chromatin	6
1.2.2 <i>Cis</i> -regulatory elements and transcription factor binding	7
1.2.3 The role of three-dimensional genome architecture in gene regulation	8
1.2.4 Insulator proteins and architectural factors	8
1.2.5 Chromatin remodelers and loop formation	9
1.2.6 Intrinsically disordered regions and transcriptional condensates in 3D genome regulation	9
1.3 Pioneer transcription factor competition in three-dimensional space	10
1.3.1 Pioneer transcription factors	10
1.3.2 CLAMP and GAF as GA-binding pioneer transcription factors	11
1.3.3 Transcription factor competition at shared binding motifs	12
1.3.4 Context-specific chromatin loops by CLAMP and GAF	13
1.3.5 Competition for DNA binding between paralogous transcription factors	14
1.4 Investigating co-regulation in the <i>omics</i> era	14
1.4.1 Genome-wide approaches to transcriptional regulation	14
1.4.2 Nascent transcription analysis	15
1.4.3 Three-dimensional genomics	15
1.4.4 Integrating multi-omic datasets	16
1.5 Towards a unified omics model of context-specific gene regulation	17
1.5.1 From descriptive integration to predictive modeling	17
1.5.2 Machine learning approaches for regulatory genomics	18
1.6 Thesis overview	19

2	Chapter 2: Differential chromatin looping regulated by two GA-binding transcription factors creates an X-specific chromatin environment for dosage compensation	21
2.1	Abstract	22
2.2	Introduction	22
2.3	Results	26
2.3.1	CLAMP- and GAF-dependent chromatin loops are largely non-overlapping and have different size distributions	26
2.3.2	CLAMP and GAF are required for different sized chromatin loops near GA-repeat rich high-affinity sites (HAS)	31
2.3.3	Single nucleotide variation within GA-rich motifs correlates with TF-specific chromatin looping at HAS	33
2.3.4	GAF regulates chromatin looping at silent euchromatin specifically on the X chromosome and not on autosomes while CLAMP regulates chromatin looping between active chromatin regions throughout the genome	39
2.3.5	Chromatin loops directly mediated by CLAMP and GAF associate with <i>roX2</i> chromatin loops on the X chromosome	43
2.4	Discussion	46
2.5	Methods	49
2.5.1	<i>Cell Culture</i>	49
2.5.2	<i>Western Blot</i>	49
2.5.3	<i>Micro-C</i>	50
2.5.4	<i>Hi-ChIP</i>	51
2.5.5	<i>Micro-C Preprocessing</i>	51
2.5.6	<i>Hi-ChIP Preprocessing</i>	51
2.5.7	<i>Computational Analysis</i>	52
2.6	Data Access	57
2.7	Competing interest statement	57
2.8	Acknowledgments	57
2.9	Author Contributions	57
2.10	Supplemental Figures	58
2.11	Supplemental Tables	62
3	Chapter 3: The CLAMP GA-binding transcription factor regulates heat stress-induced transcriptional repression	65
3.1	Abstract	66
3.2	Introduction	66
3.3	Results	69
3.3.1	CLAMP is required for the repression of transcription in response to heat stress	69
3.3.2	Heat stress induces widespread local 3D chromatin changes at chromatin loop anchors	72
3.3.3	CLAMP inhibits chromatin loop formation during HS and regulates repression of associated genes through both direct and indirect mechanisms	76
3.3.4	Specific combinations of co-factors present prior to heat stress distinguish the direct from the indirect functions of CLAMP in 3D looping and heat stress repression	81
3.4	Discussion	84
3.5	Materials and methods	87

3.5.1	Fly strains and husbandry	87
3.5.2	Cell culture	87
3.5.3	Polytene chromosome squashes and immunostaining after HS	87
3.5.4	Western Blotting to validate <i>clamp</i> dsRNA treatment for <i>clamp</i> RNAi	88
3.5.5	SLAM-seq	88
3.5.6	Micro-C	89
3.5.7	Hi-ChIP	89
3.5.8	Computational analysis	90
3.6	Competing Interest Statement	93
3.7	Funding	93
3.8	Author Contributions	93
3.9	Supplemental Figures	94
3.10	Supplementary Tables	99
4	Chapter 4: Leveraging graph neural networks to classify GA-repeat motifs using epigenomic features and 3D chromatin structure	105
4.1	Abstract	106
4.2	Introduction	106
4.3	Experimental Design and Methods	109
4.3.1	Overview of datasets	109
4.3.2	Generation of target variable	110
4.3.3	Graph construction and data integration	110
4.3.4	Baseline models	111
4.3.5	Graph convolutional neural networks	112
4.3.6	Evaluation metrics	113
4.3.7	Experimental overview	114
4.4	Results and Discussion	114
4.4.1	Graph convolutional neural networks outperform baseline models	114
4.4.2	Class-specific combinations of histone modifications emerge from classifier	115
4.4.3	Unique epigenome signatures are found between and within target classes	117
5	Chapter 5: Discussion	120
5.1	A single cytosine within GA-repeats dictates pioneer transcription factor identity and 3D chromatin architecture	121
5.2	Transcription factor competition at shared motifs offers a framework for interpreting non-coding genetic variation	122
5.3	A chromatin-based regulatory scaffold for heat stress response	124
5.4	CLAMP produces opposing transcriptional outcomes in different regulatory contexts	126
5.5	Pitfalls to consider	128
5.6	Summary	128
	References	130

List of Figures

2.1	Loss of CLAMP versus GAF alters different chromatin loops.	27
2.2	Loss of CLAMP and GAF alters chromatin loop architecture at High Affinity Sites (HAS) on the X chromosome.	32
2.3	Direct looping by CLAMP and GAF suggests size-specific loop regulation at the HAS.	34
2.4	Chromatin State Pair enrichment of CLAMP- and GAF-dependent loops shows GAF regulates chromatin looping at silent euchromatin on the X chromosome and not on autosomes.	40
2.5	Direct CLAMP- and GAF-dependent loops are associated with a subset of <i>roX2</i> chromatin loops on the X chromosome.	44
2.6	Model for X-specific CLAMP and GAF function in chromatin looping.	47
S2.1	Depletion of CLAMP or GAF in Micro-C samples	58
S2.2	Micro-C contact matrices across all <i>Drosophila melanogaster</i> chromosomes	59
S2.3	TAD organization changes across the genome upon knockdown of CLAMP and GAF	60
S2.4	The sizes of <i>roX2</i> -associated chromatin loops are significantly larger at HAS than on the X chromosome	61
S2.5	Majority of <i>roX2</i> -associated chromatin loops encompassing dosage compensated genes are at HAS on the X chromosome	62
3.1	CLAMP regulates heat-stress-induced transcriptional repression more strongly than activation	70
3.2	Heat stress induces widespread local 3D chromatin changes at the loop anchors	73
3.3	CLAMP regulates chromatin loop changes during HS, directly as well as indirectly, to repress gene transcription	76
3.4	HS-associated chromatin loops directly and indirectly regulated by CLAMP have different chromatin-bound factors under NHS conditions	82
3.5	Model for CLAMP-dependent HS-induced transcriptional gene repression	85
S3.1	Photomicrograph showing validation of CLAMP dsRNA treatment, lowering CLAMP protein level in Kc cells	94
S3.2	CLAMP functions more frequently as a repressor than an activator in both NHS and HS conditions	94
S3.3	CLAMP-dependent HS-repressed genes significantly overlap with HS-repressed genes	95
S3.4	Many CLAMP-dependent repressed genes are present at the loop anchors of CLAMP-dependent HS-associated chromatin loops lost after HS	96
S3.5	CLAMP-HiChIP identifies CLAMP-bound chromatin loops in control and HS cells	97
S3.6	Most CLAMP-associated changes upon heat stress are three-dimensional and not associated with differences in CLAMP occupancy	98
4.1	Experimental overview of the GA-repeat classification framework	113
4.2	The GCN outperforms baseline models for GA-repeat classification	115

4.3	Class-specific histone modification importance scores derived from GNNExplainer .	116
4.4	Hierarchical clustering of GNNExplainer importance scores reveals unique epigenomic signatures within and between GA-repeat classes	118
5.1	CLAMP performs context-specific regulatory roles in male and female <i>Drosophila</i> cell lines	127

List of Tables

S2.1 All detected differential Micro-C chromatin interactions	63
S2.2 Normalization of roX2, CLAMP-dependent, and GAF-dependent Loops	64
S3.1 Statistical tests.	99

Abstract

Transcriptional co-regulation, the coordinated regulation of gene sets, is essential for development, homeostasis, and environmental responsiveness. However, the precise mechanisms by which specific regions of the genome are targeted for co-regulation remain unknown. A critical question about gene regulation that has been outstanding for the last fifty years is: How are the right genes targeted at the right time for co-regulation? I have made significant progress addressing this important specificity question using two well-established model systems: dosage compensation and heat stress. The mechanism by which transcription factors achieve context-specific co-regulation remains poorly understood in these systems despite decades of research. Here, I identify specificity mechanisms by generating and integrating three-dimensional multi-omics data across two complementary regulatory systems: dosage compensation (Chapter 2) and the heat stress response (Chapter 3). I also build and test multiple classes of machine learning models to predict which epigenomic factors work combinatorially together to distinguish a specific class of *cis*-regulatory elements (Chapter 4).

In *Drosophila*, the essential GA-binding pioneer transcription factors CLAMP and GAF compete for occupancy at shared GA-repeat motifs that differ by only one SNP, yet produce distinct transcriptional outcomes in different biological contexts. In male *Drosophila* cells, where dosage compensation is constitutively active, I demonstrate that CLAMP and GAF directly regulate largely mutually exclusive 3D chromatin contacts. CLAMP drives short-range loops linking high-affinity dosage compensation complex (DCC) binding sites with active, dosage-compensated housekeeping genes on the X chromosome. In contrast, GAF mediates longer-range loops between transcriptionally silent insulator regions specifically on the X-chromosome but not on autosomes. On autosomes, CLAMP and GAF regulate looping at different subsets of active regions. These findings reveal how differential transcription factor occupancy creates an X-chromosome-specific chromatin environment for dosage compensation. In female *Drosophila* cells, which lack active dosage compensation, I identify CLAMP as the first DNA-binding transcription factor that directly regulates heat stress-induced repression of constitutive genes through modulation of 3D chromatin looping. I determined that CLAMP represses approximately 75% of heat stress-repressed genes and functions through a specific insulator-remodeler complex composed of Ibf1/2, ZIPIC, and ISWI.

My thesis work demonstrates that pioneer transcription factor competition at shared binding motifs drives context-specific co-regulation through differential modulation of 3D chromatin architecture. Finally, I developed interpretable graph neural networks for our integrated multi-omic datasets to computationally dissect the regulatory logic underlying context-specific co-regulation (Chapter 4). My models identified a novel link between histone modifications and dosage compensation which generates many new hypotheses to test. Together, I have discovered new three-dimensional mechanisms by which specific genes are co-regulated over space and time through transcription factor competition. Many SNPs linked to disease are located within transcription factor binding sites, and my thesis work identifies new rules for co-regulation of genes that will allow us to predict how SNPs cause transcriptional dysregulation in the future and how to correct these defects.

Chapter 1: Introduction

Joseph Aguilera¹ and Erica Larschan¹

I wrote this chapter with editorial feedback from Dr. Erica Larschan.

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

1.1 Co-regulation of the *Drosophila melanogaster* genome

1.1.1 What is transcriptional co-regulation and why does it matter?

Within cells, the coordinated regulation of specific groups of genes, referred to as transcriptional co-regulation, is essential for development, homeostasis, and environmental responsiveness (Lee and Young 2013). Rather than regulating each gene independently, cells activate or repress functionally related gene sets in a temporally and spatially manner to facilitate specific transcriptional programs (Lee and Young 2013). Transcription factors (TFs) recruit large transcription complexes to facilitate spatially precise and temporally dynamic transcriptional repression and activation of specific groups of genes (Zaret 2020; Drouin 2014). Therefore, understanding the mechanisms by which TFs achieve coordinate regulation of gene sets is a fundamental question in molecular biology.

Co-regulatory programs can be broadly classified based on the type of the stimulus that elicits them. *Endogenous* co-regulation encompasses developmental programs and constitutive cellular processes, such as the sex-specific equalization of X-linked gene dosage between males and females, known as dosage compensation (Lucchesi and Kuroda 2015; Samata and Akhtar 2018). *Exogenous* co-regulation arises in response to environmental perturbations—such as heat stress, oxidative damage, or pathogen exposure—that require rapid genome-wide transcriptional reprogramming to ensure cellular survival (Vihervaara et al. 2018). While these two classes of co-regulation differ in their temporal dynamics and biological purpose, they share a common mechanistic challenge: *how does the cell select the correct subset of genes for coordinate regulation from among the thousands encoded in the genome?*

1.1.2 *Drosophila melanogaster* as a model for studying co-regulation

Drosophila melanogaster (the fruit fly) has served as a powerful genetic model organism for over a century, and its utility for the study of transcriptional regulation is unmatched among metazoans. The *Drosophila* genome is compact (~180 Mb) yet encodes approximately 14,000 protein-coding genes, many of which are conserved in humans (Adams et al. 2000). Importantly, *Drosophila* has excellent tools for dissecting transcriptional mechanisms, which make many conserved transcrip-

tional programs best studied in this system (Samata and Akhtar 2018).

Several features of the *Drosophila* system make it especially well-suited for the study of co-regulation. First, a wealth of genome-wide datasets—including chromatin state maps, TF binding profiles, and chromatin accessibility measurements—have been generated through large-scale consortia such as modENCODE (Gerstein et al. 2010) and are publicly available through FlyBase (Gramates et al. 2022). Second, powerful genetic tools enable precise loss-of-function analyses. For example, RNA interference (RNAi) permits efficient and inducible knockdown of individual genes in cultured *Drosophila* cell and fly lines, and most recently, CRISPR-based approaches allow targeted mutagenesis in vivo (Gratz et al. 2013). Third, the larval salivary glands of *Drosophila* produce polytene chromosomes that permit direct cytological visualization of TF binding and gene activity at defined chromosomal positions (Zhimulev et al. 2004). Together, these resources establish *Drosophila* as an ideal system to dissect the mechanisms of context-specific co-regulation at a high resolution.

1.1.3 Two paradigms of precise co-regulation in *Drosophila*

There are many biological contexts where precise co-regulation is required, including dosage compensation and the heat stress response. While these processes are fundamentally distinct—one is an endogenous, constitutive, sex-specific program and the other is an exogenous, inducible, stress-responsive program—both require the coordinate regulation of hundreds to thousands of genes. Furthermore, both are mediated in part by the same class of GA-binding pioneer TFs acting through the 3D chromatin architecture.

Dosage compensation

In organisms with heteromorphic sex chromosomes, the difference in X chromosome copy number between males (XY) and females (XX) creates an inherent imbalance in the dosage of X-linked genes. Dosage compensation mechanisms have evolved independently across species to correct this imbalance. In *Drosophila*, dosage compensation is achieved by the approximately two-fold transcriptional upregulation of genes on the single male X chromosome, a process known as X chromosome upregulation (XCU) (Lucchesi and Kuroda 2015). XCU, which equalizes the levels of X-chromosome and autosomal gene expression within heterogametic species, is the most highly

conserved step in dosage compensation across species. Importantly, this step maintains the stoichiometry of the many protein complexes with components encoded on both the X chromosome and autosomes (Samata and Akhtar 2018). Recent work has identified convergent evolution toward the enrichment of the H4K16ac chromatin mark which increases transcript levels two-fold as the most highly conserved mechanism for dosage compensation which occurs across heterogametic species (Zimmer et al. 2016; Gu et al. 2019).

XCU is mediated by the Male-Specific Lethal (MSL) complex, also known as the dosage compensation complex (DCC). The DCC is composed of five protein subunits (MSL1, MSL2, MSL3, MOF, and MLE) and two long non-coding RNAs, roX1 and roX2 (Samata and Akhtar 2018). The DCC is recruited to the X chromosome through a hierarchical process that begins at the high-affinity sites (HAS), also referred to as chromatin entry sites (CES), which are enriched for GA-repeat motifs (Alekseyenko et al. 2008; Kelley et al. 1999; Straub et al. 2008). Subsequent to identifying these binding sites that are already occupied by CLAMP in early development (Rieder et al. 2017), the DCC spreads along the length of the X chromosome through the function of DCC components, including the roX non-coding RNAs (Rieder et al. 2019; Valsecchi et al. 2018; Park et al. 2002; Kelley et al. 1999). The histone acetyltransferase MOF catalyzes acetylation of histone H4 at lysine 16 (H4K16ac), a modification associated with transcriptional upregulation (Gelbart and Kuroda 2009).

A central question in the field for decades has been how the DCC distinguishes the X chromosome from autosomes. The GA-repeat motifs at HAS are necessary but not sufficient for DCC targeting, as similar GA-repeat motifs exist throughout the autosomal genome (Kuzu et al. 2016). As we discuss in Chapter 2, the differential occupancy of GA-binding TFs at these sites provides a critical mechanism for X-specific DCC recruitment. The male *Drosophila* S2 cell line, in which the dosage compensation machinery is constitutively active, provides an ideal experimental system for studying the interplay between TF competition and DCC targeting in the context of XCU.

Heat stress response

The heat stress response is among the most conserved transcriptional programs in biology, enabling organisms to survive acute exposure to elevated temperatures. In *Drosophila*, as in other eukaryotes, heat stress triggers the rapid activation of heat shock genes, most prominently the

Hsp70 family. This response is mediated by Heat Shock Factor (HSF), which binds to heat shock elements (HSEs) in target gene promoters upon heat stress (Vihervaara et al. 2018). GAF is one of the first TFs to be identified across species and recognizes GA-repeat motifs throughout the genome to mediate activation of transcription in many contexts including regulation of the heat shock genes (Duarte et al. 2016; Adkins et al. 2006).

While the activation role of the heat stress response has been extensively characterized, a less studied but equally critical component is the global repression of constitutive transcription that accompanies heat shock gene induction. Upon heat stress, the vast majority of actively transcribed genes are rapidly silenced, a phenomenon that has been documented cytologically on polytene chromosomes and RNA sequencing (Mahat et al. 2016). This transcriptional repression is thought to redirect the transcriptional machinery toward stress-responsive loci, but the molecular mechanism by which this widespread repression is achieved has remained largely unknown.

Our recent work has identified CLAMP (Chromatin-Linked Adaptor for MSL Proteins) as the first reported DNA-binding factor that directly regulates the repression of constitutive genes upon heat stress, mediating approximately 75% of heat stress-induced transcriptional repression through modulation of 3D chromatin looping (Aguilera et al. 2025). Importantly, these experiments were performed in the *Drosophila* female Kc167 cell line, which lacks an active dosage compensation complex. The use of Kc167 cells is a key experimental choice as these cells do not express the DCC. Therefore, CLAMP's role in heat stress-mediated transcriptional repression can be studied independently of its function in dosage compensation, thereby cleanly isolating the exogenous regulatory context from the endogenous one.

1.1.4 GA-binding pioneer transcription factors as central regulators

The two co-regulatory systems described above, dosage compensation and the heat stress response, converge on a shared class of regulatory proteins: the GA-binding pioneer TFs CLAMP and GAF (GAGA Factor, encoded by the *Trithorax-like* gene). Both factors recognize GA-repeat sequences that are distributed throughout the *Drosophila* genome and are enriched at promoters, enhancers, and HAS for the DCC (Fuda et al. 2015; Kuzu et al. 2016). Variation within GA-repeat binding sites that accumulated over evolutionary time on the X chromosome promotes the occupancy of CLAMP instead of GAF (Kaye et al. 2017; Kuzu et al. 2016).

Apart from dosage compensation, these two TFs have opposing transcriptional functions at most genomic loci. GAF often acts as a transcriptional activator (Adkins et al. 2006; Duarte et al. 2016), while CLAMP often functions as a repressor (Urban et al. 2017). Furthermore, CLAMP is a pioneer TF that binds to GA-repeat motifs to regulate chromatin accessibility and remodeling in other contexts, such as zygotic genome activation (Duan et al. 2021) and histone gene transcription (Rieder et al. 2017). Variation in sequence composition at GA-repeat motifs determines whether CLAMP or GAF will occupy a specific site, with CLAMP but not GAF tolerating variation in GA-repeats that appear within the strongest of DCC binding sites (HAS) (Kaye et al. 2017; Adkins et al. 2006). While CLAMP but not GAF directly associates with the DCC both *in vivo* and *in vitro* (Alekseyenko et al. 2015; Eggers et al. 2023), GAF promotes DCC recruitment but to a much lesser extent (Kaye et al. 2017). Therefore, understanding the competitive relationship between GAF and CLAMP will also reveal new insight into how the balance between activating and repressive transcriptional programs is established across the genome.

This thesis addresses a central question that emerges from this biology: how do two pioneer TFs with overlapping binding specificities achieve context-specific co-regulation of distinct gene sets? By examining CLAMP and GAF competition in the context of dosage compensation (in male S2 cells) and CLAMP-mediated repression in the context of heat stress (in female Kc167 cells), our work demonstrates that the same molecular players—CLAMP, GAF, and their associated insulator and chromatin remodeling co-factors—are deployed in both regulatory contexts. However, these shared factors produce fundamentally distinct transcriptional outcomes through differential modulation of 3D chromatin architecture. Together, these complementary systems reveal how small differences in DNA sequence and chromatin environment determine the context-specific recruitment of large transcription complexes that induce coordinated regulation of specific groups of genes.

1.2 Chromatin organization and transcriptional regulation

1.2.1 Hierarchical organization of eukaryotic chromatin

Within the nucleus of eukaryotic cells, chromatin is tightly compacted through multiple levels of organization (Bonev and Cavalli 2016) and shares finite nuclear space with many classes of tran-

scriptional regulators, including DNA-binding TFs and large transcription complexes which are often targeted to chromatin by TFs (Zaret 2020). At the most fundamental level, approximately 147 base pairs of DNA are wrapped around an octamer of histone proteins to form the nucleosome, the basic repeating unit of chromatin (Luger et al. 1997). Nucleosomes are further organized into higher-order structures including chromatin loops, topologically associating domains (TADs), compartments, and chromosome territories (Luger et al. 1997; Dixon et al. 2012). Notably, these structures are not static; they are dynamically regulated in response to developmental signals, environmental stimuli, and the cell cycle, and their disruption can lead to aberrant gene expression and disease (Bonev and Cavalli 2016).

In *Drosophila*, TFs and transcription complexes play an important role facilitating chromatin loops that bring together specific groups of genes for organization and transcription through the function of insulators and tethering elements (Gutierrez-Perez et al. 2019; Boettiger et al. 2016; Melnikova et al. 2004; Wolle et al. 2015; Kaye et al. 2018; Levo et al. 2022; Varisco et al. 2025; Li et al. 2023). Furthermore, recent work indicates that loop extrusion mechanisms function together with tethering elements containing GA-rich sequences to promote loop formation and transcriptional bursting (Choppakatla et al. 2026). While loop extrusion increases the rate of the search for the correct 3D contact, tethering elements are essential to stabilize these contacts and promote transcriptional bursting through heterotypic interactions between different TFs (Varisco et al. 2025; Choppakatla et al. 2026). However, the mechanisms by which specific TFs bring together the correct target loci in a context-specific manner to precisely co-regulate genes remain poorly understood (Abidi et al. 2026).

1.2.2 *Cis*-regulatory elements and transcription factor binding

A major component of gene expression is driven by *cis*-regulatory elements—including promoters, enhancers, silencers, and insulators—that serve as platforms for TF binding and integration of regulatory information. Promoters direct the assembly of the pre-initiation complex and position RNA polymerase II (Pol II) for transcription initiation, while enhancers recruit activating TFs and co-activators that can stimulate transcription from distances of tens to hundreds of kilobases (Spitz and Furlong 2012). Silencers recruit repressive factors that inhibit transcription, and insulators function as boundary elements that prevent inappropriate regulatory interactions between

neighboring chromatin domains (Kyrchanova and Georgiev 2014). The combinatorial binding of TFs at these *cis*-regulatory elements determines the precise spatiotemporal expression pattern of each gene, and variation within these sequences can alter TF occupancy to drive context-specific transcriptional outcomes (Spitz and Furlong 2012; Kuzu et al. 2016).

1.2.3 The role of three-dimensional genome architecture in gene regulation

The 3D organization of chromatin within the nucleus is not merely a consequence of gene activity but an active driver of transcriptional regulation. Chromatin looping brings distal regulatory elements into physical proximity with their target promoters, enabling regulatory interactions that would not be possible based on linear genomic distance alone (Dekker and Mirny 2016). At a higher level of organization, the genome is partitioned into TADs, self-interacting chromatin domains that constrain regulatory interactions and are largely conserved across cell types (Dixon et al. 2012). In *Drosophila*, prior low-resolution studies suggested that the strongest DCC binding sites cluster in 3D space and CLAMP contributes to this process (Jordan and Larschan 2021). GAF has also been shown to contribute to 3D chromatin looping in the nervous system (Li et al. 2023). Therefore, understanding how TFs regulate chromatin looping is essential for understanding context-specific co-regulation.

1.2.4 Insulator proteins and architectural factors

In mammals, the zinc finger protein CTCF is the primary insulator protein and plays a central role in defining chromatin domain boundaries and mediating long-range chromatin interactions with the cohesin complex (Merkenschlager and Nora 2016). TFs such as CTCF are important for defining boundaries between 3D domains by regulating the loop extrusion process among other functions (Nora et al. 2017). In mammals, additional factors that bind to similar binding sites function redundantly with CTCF but their relationship with CTCF remains poorly understood (Merkenschlager and Odom 2013). In *Drosophila*, the insulator landscape is more complex, with multiple insulator proteins including CP190, BEAF-32, Su(Hw), Ibf1/2, and ZIPIC contributing to chromatin domain organization (Kyrchanova and Georgiev 2014). CP190 functions as a central hub that interacts with many DNA-binding insulator proteins, while BEAF-32 preferentially associates with active promoter regions (Jiang et al. 2009). Ibf1 and Ibf2 are CP190-interacting insulator binding factors that con-

tribute to chromatin boundary function (Cuartero et al. 2014). ZIPIC is a zinc finger protein that promotes CP190-mediated insulator activity (Maksimenko et al. 2015). These insulator proteins are particularly notable because Ibf1/2, ZIPIC, and the chromatin remodeler ISWI are significantly enriched at CLAMP-dependent chromatin loops in both dosage compensation and the heat stress response (Aguilera et al. 2025; Aguilera et al. 2026).

1.2.5 Chromatin remodelers and loop formation

ATP-dependent chromatin remodelers alter the position, composition, and spacing of nucleosomes, thereby modulating the accessibility of DNA to TFs and other regulatory factors (Clapier et al. 2017). Among these, the ISWI family of remodelers is particularly well-characterized in *Drosophila*, where ISWI-containing complexes such as NURF, ACF, and CHRAC regulate nucleosome positioning at promoters, enhancers, and chromatin boundaries (Tsukiyama et al. 1995; Ito et al. 1997). ISWI has also been implicated in higher-order chromatin organization; loss of ISWI leads to global disruption of chromosome structure visible on polytene chromosomes (Deuring et al. 2000). The enrichment of ISWI at CLAMP-regulated chromatin loops in both dosage compensation and the heat stress response suggests that CLAMP may recruit or cooperate with ISWI-containing complexes to modulate local nucleosome positioning and thereby facilitate or disrupt specific chromatin loops (Aguilera et al. 2025; Aguilera et al. 2026). This connection between chromatin remodeling and 3D chromatin loop formation represents a key mechanistic link between TF binding and the organization of the 3D genome.

1.2.6 Intrinsically disordered regions and transcriptional condensates in 3D genome regulation

Many transcription factors contain intrinsically disordered low-complexity domains (LCDs) that mediate dynamic protein-protein interactions, driving the formation of local high-concentration transcription factor hubs at target genomic loci (Sabari et al. 2018; Boija et al. 2018). These hubs recruit RNA Pol II, Mediator, and transcriptional cofactors, and their assembly is tightly regulated (Yoshida et al. 2026). Importantly, hub formation behavior is sensitive to cellular context as endogenous hubs are maintained by active dissolution mechanisms that prevent macroscopic phase separation (Yoshida et al. 2026). In the context of stress, transcription factor condensates can

be rapidly induced to reorganize the 3D genome; upon heat stress, HSF1 forms transcriptional condensates that drive extensive inter- and intra-chromosomal interactions between target loci, coupling phase separation to chromatin looping and transcriptional activation (Chowdhary et al. 2022; Mohajan and Gross 2025). Furthermore, IDR-containing transcription factors at enhancers can directly link condensate nucleation to 3D genome architecture, establishing a mechanistic connection between TF binding, phase separation, and chromatin loop formation (Tiusanen et al. 2026). These observations raise the possibility that heat stress may alter the phase transition properties of CLAMP, GAF, or their associated co-factors, contributing to the context-dependent chromatin looping and transcriptional outcomes described in this work.

1.3 Pioneer transcription factor competition in three-dimensional space

1.3.1 Pioneer transcription factors

A subset of transcription factors, termed pioneer factors, possess the unique ability to engage condensed or nucleosome-occupied chromatin and initiate regulatory events at previously inaccessible genomic loci (Zaret 2020). Unlike conventional TFs, which require pre-existing chromatin accessibility to bind their target sequences, pioneer factors directly contact nucleosomal DNA through specialized DNA-binding domains that tolerate the steric constraints imposed by histone octamers (Zaret and Carroll 2011). Upon binding, pioneer factors recruit chromatin remodeling complexes and histone-modifying enzymes that locally displace or reposition nucleosomes, thereby creating accessible chromatin states that permit the subsequent binding of additional TFs and the assembly of active transcriptional machinery (Cirillo et al. 2002; Iwafuchi-Doi and Zaret 2014). This hierarchical binding model, in which a pioneer factor initiates a cascade of chromatin remodeling and TF recruitment, is now recognized as a fundamental mechanism for establishing new regulatory states during development, differentiation, and cellular reprogramming (Zaret 2020).

The functional significance of pioneer factor activity extends beyond simple chromatin opening. Pioneer factors serve as molecular scaffolds that recruit co-factors, establish chromatin domains, and in many cases, define the competence of a genomic locus to respond to subsequent signaling inputs (Drouin 2014; Larson et al. 2021). In *Drosophila*, two well-characterized pioneer factors, CLAMP and GAF, bind GA-repeats distributed throughout the genome and regulate chromatin

accessibility at thousands of genomic loci (Duan et al. 2021; Fuda et al. 2015). Importantly, these two factors recognize highly similar DNA sequence motifs yet mediate fundamentally distinct transcriptional outcomes depending on the regulatory context, which raises the question of how two pioneer factors with overlapping binding specificities achieve context-specific gene regulation.

1.3.2 CLAMP and GAF as GA-binding pioneer transcription factors

CLAMP (Chromatin-Linked Adaptor for MSL Proteins) was originally identified through its essential role in recruiting the dosage compensation complex to the male X chromosome in *Drosophila* (Larschan et al. 2012; Soruco et al. 2013). Subsequent work established CLAMP as an essential pioneer transcription factor; CLAMP binds to genome-wide GA-repeats to regulate chromatin accessibility and remodeling in diverse contexts including zygotic genome activation and histone gene transcription (Duan et al. 2021; Rieder et al. 2017). CLAMP is encoded by a single gene, is maternally deposited, and is essential for viability in both sexes, indicating that its functions extend well beyond the male-specific process of dosage compensation (Rieder et al. 2017; Alekseyenko et al. 2008; Alekseyenko et al. 2012). Structurally, CLAMP contains multiple C2H2 zinc finger domains that mediate DNA binding with a preference for GA-repeat sequences containing cytosine interruptions, a feature that distinguishes its binding specificity from that of GAF (Kaye et al. 2017).

GAF (GAGA Factor), encoded by the *Trithorax-like* (Trl) gene, is one of the first pioneer transcription factors identified across species and recognizes GA-repeat motifs throughout the genome to mediate activation of transcription in many contexts, including zygotic genome activation and regulation of the heat shock genes (Gaskill et al. 2021; Duarte et al. 2016; Adkins et al. 2006). GAF plays critical roles in promoter-proximal pausing of RNA Polymerase II, nucleosome remodeling, enhancer function (Fuda et al. 2015), and it functions as a chromatin insulator component that facilitates long-range interactions between regulatory elements (Li et al. 2023; Wolle et al. 2015; Gaskill et al. 2021). Similar to CLAMP, GAF is ubiquitously expressed and essential for viability, underscoring the broad importance of GA-binding pioneer factors in *Drosophila* gene regulation.

Despite their shared GA-repeat binding specificity, CLAMP and GAF have distinct protein architectures and interact with different sets of co-factors, resulting in divergent functional roles at many genomic loci (Kaye et al. 2018). GAF often acts as a transcriptional activator, whereas CLAMP often functions as a repressor at most genomic loci (Adkins et al. 2006; Duarte et al. 2016; Urban

et al. 2017). Critically, CLAMP but not GAF directly associates with the DCC both *in vivo* and *in vitro* (Alekseyenko et al. 2015; Eggers et al. 2023), while GAF promotes DCC recruitment but to a much lesser extent (Kaye et al. 2018). Furthermore, CLAMP interacts with the Negative Elongation Factor (NELF), which functions in heat shock repression (Aguilera et al. 2025), whereas GAF and HSF are involved in heat-stress-induced activation but not repression (Duarte et al. 2016). These distinct co-factor interactions underlie the context-dependent synergistic and antagonistic relationships between CLAMP and GAF (Kaye et al. 2018).

1.3.3 Transcription factor competition at shared binding motifs

Many transcription factors that are members of similar functional classes occupy similar binding sites, and variation within these sites can drive differential occupancy (Zaret 2020; Jaeger et al. 2010; Kuzu et al. 2016). In *Drosophila*, CLAMP and GAF compete for occupancy at shared GA-repeat motifs, and the outcome of this competition is determined by subtle sequence variation within seemingly similar cis-elements (Kaye et al. 2018; Kuzu et al. 2016). Specifically, variation in sequence composition at GA-repeat motifs determines whether CLAMP or GAF will occupy a specific site. CLAMP tolerates single cytosine nucleotide interruptions within stretches of GA-repeats, while GAF preferentially binds contiguous GA-repeat motifs that lack such interruptions (Adkins et al. 2006; Kaye et al. 2018). This molecular-level specificity is particularly evident at the strongest DCC binding sites (HAS/CES) on the X chromosome, where sequence variation that accumulated over evolutionary time promotes CLAMP binding while excluding GAF (Kaye et al. 2018; Kuzu et al. 2016).

The competitive relationship between CLAMP and GAF extends beyond simple binding site occupancy. Depletion of either factor leads to reciprocal changes in occupancy of the other at shared sites, demonstrating that the balance between these two pioneer factors is dynamically maintained (Kaye et al. 2018). Furthermore, the functional consequence of this competition differs between the X chromosome and autosomes; on the X chromosome, CLAMP outcompetes GAF at HAS and recruits the DCC, while on autosomes both factors co-occupy GA-rich sites without X-specific complex recruitment (Eggers and Becker 2021; Eggers et al. 2023). This chromosome-specific outcome of TF competition represents a powerful mechanism for achieving context-specific co-regulation from a shared set of DNA-binding proteins.

1.3.4 Context-specific chromatin loops by CLAMP and GAF

Prior low-resolution studies suggested that the strongest DCC binding sites cluster in 3D space and that CLAMP promotes long-range chromatin interactions (Jordan and Larschan 2021). GAF has also been shown to contribute to 3D chromatin looping in the nervous system (Li et al. 2023). However, the mechanisms by which these two factors with similar binding specificities differentially regulate 3D chromatin loop formation had not been directly addressed. Our recent work has revealed that CLAMP and GAF mediate largely mutually exclusive 3D chromatin loops, creating distinct chromosome-specific chromatin environments through differential chromatin looping (Aguilera et al. 2026).

In male S2 cells, where dosage compensation is constitutively active, CLAMP drives local short-range interactions that directly link high-affinity DCC binding sites with active, dosage-compensated housekeeping genes on the X chromosome (Aguilera et al. 2026). In contrast, GAF mediates longer-range interactions between transcriptionally silent insulator regions on the X chromosome spanning a wider range of genomic distances. On autosomes, both CLAMP and GAF directly regulate contacts between active chromatin regions including active promoters, but the X-chromosome-specific short-range looping by CLAMP is absent, consistent with the lack of DCC recruitment to autosomes. These findings demonstrate that differential occupancy of CLAMP versus GAF at GA-repeat motifs on the X chromosome drives the formation of a unique chromatin architecture that facilitates dosage compensation.

In the context of the heat stress response, CLAMP plays an equally critical but mechanistically distinct role in regulating 3D chromatin architecture. In female Kc167 cells, which lack active dosage compensation, heat stress induces widespread changes in local 3D chromatin looping that are significantly associated with heat stress-induced transcriptional changes (Aguilera et al. 2025). Strikingly, CLAMP inhibits the formation of 3D chromatin loops during heat stress at loci where it is directly bound, and this inhibition of looping is associated with the repression of approximately 75% of all heat stress-repressed genes (Aguilera et al. 2025). At these direct CLAMP-regulated sites, the insulator proteins Ibf1/2 and ZIPIC and the chromatin remodeler ISWI are enriched prior to heat stress, suggesting that CLAMP functions within a specific insulator-remodeler complex to modulate loop architecture upon heat stress (Aguilera et al. 2025). These findings represent the

first identification of a DNA-binding transcription factor that directly regulates the repressive role of the heat stress response through modulation of 3D chromatin interactions.

1.3.5 Competition for DNA binding between paralogous transcription factors

Direct competition for DNA binding between paralogous transcription factors can be a major determinant of their genomic occupancy and regulatory output. Using a high-throughput *in vitro* competition PBM assay paired with *in vivo* ChIP-seq analysis in *S. cerevisiae*, Zhang et al. (2021) demonstrated that Cbf1 and Pho4 transcription factors compete depending on their relative concentrations and intrinsic divergence in DNA-binding affinity. A simple biophysical model incorporating paralog-specific dissociation constants accurately predicted the competitive binding patterns observed both *in vitro* and *in vivo*, and this competition was sufficient to shape the differential activation of Pho4 target genes in the phosphate-signaling pathway. These findings establish that even modest divergence in binding specificity between co-expressed paralogs can have substantial consequences for gene regulation. In addition, this work provides a framework directly relevant to understanding how competition between CLAMP and GAF at similar GA-repeat motifs may drive context-dependent transcriptional regulation in *Drosophila* (Zhang et al. 2021).

1.4 Investigating co-regulation in the *omics* era

1.4.1 Genome-wide approaches to transcriptional regulation

The advent of next generation sequencing technologies has fundamentally transformed our ability to dissect transcriptional regulatory mechanisms. RNA-seq enables comprehensive measurement of gene expression across the entire transcriptome (Wang et al. 2009), while chromatin immunoprecipitation followed by sequencing (ChIP-seq) and its more recent derivatives CUT&RUN and CUT&Tag permit genome-wide mapping of transcription factor binding and histone modifications with progressively lower input requirements (Park 2009; Skene and Henikoff 2017; Kaya-Okur et al. 2019). Complementary approaches for measuring chromatin accessibility, including ATAC-seq and MNase-seq, reveal the regulatory landscape of open chromatin regions genome-wide (Buenrostro JD et al. 2013; Schones et al. 2008). Together, these technologies have enabled the systematic characterization of the regulatory elements, TF binding, and chromatin states that

underlie transcriptional programs in *Drosophila* and other model organisms (Gerstein et al. 2010).

1.4.2 Nascent transcription analysis

While steady-state RNA-seq provides a snapshot of accumulated mRNA levels, it cannot distinguish between changes in transcription rate and changes in mRNA stability or processing. Nascent RNA approaches overcome this limitation by measuring RNA molecules as they are being synthesized, providing a direct readout of transcriptional activity. Techniques such as PRO-seq and GRO-seq capture elongating RNA Polymerase II and its associated nascent transcripts genome-wide (Core et al. 2008; Kwak et al. 2013). SLAM-seq (thiol(SH)-linked alkylation for the metabolic sequencing of RNA) offers an orthogonal approach by incorporating 4-thiouridine (s4U) into newly synthesized RNA, followed by chemical conversion that introduces T-to-C mutations during sequencing, thereby distinguishing nascent from pre-existing transcripts without the need for physical separation (Herzog et al. 2017).

The distinction between nascent and steady-state measurements is particularly critical for studying rapid transcriptional responses, such as the heat stress response, where mRNA stability confounds steady-state measurements and where repression of constitutive genes occurs within minutes of temperature shift (Mahat et al. 2016; Vihervaara et al. 2018). SLAM-seq has been used to define the direct transcriptional changes induced by heat stress and by CLAMP depletion in *Drosophila* Kc167 cells, enabling the identification of CLAMP as the first reported transcription factor that directly regulates the repressive role of the heat stress response (Aguilera et al. 2025).

1.4.3 Three-dimensional genomics

The development of chromosome conformation capture (3C) technologies has enabled the systematic mapping of 3D chromatin contacts that organize the genome within the nucleus (Dekker et al. 2002). The original 3C technique measured the interaction frequency between two specific genomic loci, while subsequent innovations—such as 4C (one-to-all), 5C (many-to-many), and Hi-C (all-to-all)—progressively increased the throughput and genomic coverage of chromatin interaction mapping (Simonis et al. 2006; Dostie et al. 2006; Lieberman-aiden et al. 2009). Hi-C, which combines proximity ligation with high-throughput sequencing to capture chromatin contacts genome-wide, has been instrumental in revealing the hierarchical organization of the genome into

compartments, TADs, and chromatin loops (Dixon et al. 2012; Lieberman-aiden et al. 2009).

Despite these advances, restriction enzyme-based Hi-C is limited in resolution, which does not resolve individual regulatory element interactions. Micro-C overcomes this limitation by replacing restriction enzymes with micrococcal nuclease (MNase) to fragment chromatin at the level of individual nucleosomes, achieving resolution of approximately 200 bp (Hsieh et al. 2015; Hsieh et al. 2016; Maziak et al. 2026). This nucleosome-level resolution is essential for resolving the short-range chromatin loops that link regulatory elements, such as HAS/CES to active dosage-compensated genes, which span distances that require a high resolution to detect (Aguilera et al. 2026). Furthermore, by performing differential Micro-C analyses at multiple, high resolutions and integrating the resulting contact maps, we may capture both fine-scale local interactions and broader-range chromatin contacts within a single experimental framework.

Importantly, while Micro-C reveals all chromatin contacts genome-wide, it does not directly identify which proteins mediate specific interactions. Hi-ChIP addresses this gap by combining chromatin immunoprecipitation with proximity ligation, thereby enriching for chromatin contacts that are anchored by a specific protein of interest (Mumbach et al. 2016). Additionally, by adapting the Hi-ChIP approach with MNase-based fragmentation rather than restriction enzyme digestion, a higher resolution may be achieved. CLAMP and GAF Hi-ChIP has been performed in *Drosophila* S2 and Kc167 cells to directly identify chromatin loops anchored by each factor (Aguilera et al. 2025; Aguilera et al. 2026). By intersecting Hi-ChIP data with differential Micro-C loops identified after factor depletion, we may distinguish direct from indirect regulatory effects, revealing that CLAMP and GAF mediate largely non-overlapping sets of 3D chromatin loops. Similarly, in Kc167 cells, CLAMP Hi-ChIP has identified the subset of heat stress-responsive loops that are directly bound by CLAMP and defined the co-factors enriched at these direct regulatory sites (Aguilera et al. 2025).

1.4.4 Integrating multi-omic datasets

A key challenge in modern genomics is integrating data from multiple experimental modalities to build a coherent picture of gene regulation. Each individual assay provides a partial view: ChIP-seq reveals where a factor binds but not what it does; RNA-seq measures transcriptional output but not the mechanism; and Micro-C captures 3D chromatin loops but not the factors that mediate

them. The power of multi-omic integration lies in overlaying these complementary datasets to define mechanistic relationships that no single assay can reveal. For example, integrating Micro-C, Hi-ChIP, SLAM-seq, chromatin state annotations from the modENCODE 9-state chromatin model (Kharchenko et al. 2011), and ChIP-seq binding data enables the classification of regulatory mechanisms as direct (factor-bound loop anchors associated with transcriptional changes) or indirect (loop anchors without direct factor binding associated with transcriptional changes). This integrative framework has been applied to distinguish the short-range CLAMP-dependent loops at active dosage-compensated regions from the longer-range GAF-dependent loops at insulator regions on the X chromosome (Aguilera et al. 2026), and to identify the specific insulator-remodeler complex at direct CLAMP-regulated heat stress-repressed genes (Aguilera et al. 2025).

1.5 Towards a unified omics model of context-specific gene regulation

1.5.1 From descriptive integration to predictive modeling

The integrative analyses described above define mechanistic relationships by overlaying complementary datasets, but they rely on manual curation. For example, a chromatin loop anchor must overlap both a TF binding peak and a differentially expressed gene to classify a regulatory event as direct. While effective, this approach does not scale to the full combinatorial complexity of the regulatory landscape, nor does it capture the nonlinear relationships between chromatin features that may drive TF occupancy. A complementary approach is to use machine learning to learn these relationships directly from the data, building predictive models that can identify which combinations of features are most informative for distinguishing between regulatory contexts. The CLAMP/GAF system is particularly well suited for this approach because the competition between two pioneer factors at shared GA-repeat motifs provides a clearly defined classification task with biologically interpretable outcomes.

1.5.2 Machine learning approaches for regulatory genomics

Machine learning methods have emerged as powerful tools for building predictive models of gene regulation from multi-omic data. Approaches, ranging from ensemble methods to neural networks, have been applied to predict gene expression from DNA sequence, chromatin features, and 3D contact information (Avsec et al. 2021; Kelley et al. 2018; Karbalayghareh et al. 2022). In the context of *Drosophila* regulatory genomics, the availability of comprehensive chromatin state annotations from the modENCODE project (Kharchenko et al. 2011) and the wealth of TF binding data provide a rich feature space for training such models. Furthermore, recent advances in graph neural networks for 3D genomic data have demonstrated that chromatin contact information can significantly improve the prediction of gene expression and regulatory interactions (Karbalayghareh et al. 2022), making these approaches well suited for leveraging the Micro-C and Hi-ChIP data generated in this thesis.

Building on these advances, GC-MERGE demonstrated that graph convolutional networks can effectively integrate Hi-C-derived chromatin interaction maps with histone modification profiles to predict gene expression across human cell lines, providing an interpretable framework that quantifies both node-level features and edge-level interactions in the context of 3D genome organization (Bigness et al. 2022). In this thesis, we extended this graph-based framework to the CLAMP/GAF system by constructing a genome-wide graph in which each 1kb genomic bin is a node with histone modification features and edges represent statistically significant Micro-C chromatin interactions. Using this representation, I trained a graph convolutional neural network (GCN) to classify GA-repeat motifs into biologically meaningful categories and compared its performance against traditional machine learning and deep learning baselines that lack access to 3D structural information. The complete methodology and results are presented in Chapter 4, which also lays the groundwork for future models that incorporate edge importance to reveal the spatial, regulatory logic of pioneer transcription factor occupancy.

1.6 Thesis overview

This thesis addresses the fundamental question of how pioneer transcription factors achieve context-specific co-regulation of gene expression through modulation of 3D chromatin architecture. I use the *Drosophila* GA-binding pioneer transcription factors CLAMP and GAF as a model system in which two factors with similar DNA-binding specificities compete for occupancy at shared motifs yet produce distinct regulatory outcomes depending on the biological context. By integrating Micro-C, Hi-ChIP, SLAM-seq, and chromatin state data across two complementary regulatory contexts, dosage compensation and the heat stress response, I define the mechanisms by which differential TF binding drives context-specific 3D chromatin looping and transcriptional co-regulation.

In Chapter 2, I investigate how CLAMP and GAF differentially regulate 3D chromatin loop formation on the X chromosome and autosomes in male *Drosophila* S2 cells, where dosage compensation is constitutively active. Using Micro-C and Hi-ChIP, I demonstrate that CLAMP and GAF mediate largely mutually exclusive chromatin contacts; CLAMP drives short-range loops linking high-affinity DCC binding sites to active dosage-compensated genes, while GAF mediates longer-range loops between transcriptionally silent insulator regions. These findings reveal how differential TF occupancy at shared GA-repeat motifs creates an X-chromosome-specific chromatin environment for dosage compensation.

In Chapter 3, I define the role of CLAMP in heat stress-induced transcriptional repression using female *Drosophila* Kc167 cells, where dosage compensation is absent. Through SLAM-seq, Hi-ChIP, and Micro-C, I demonstrate that CLAMP regulates approximately 75% of all heat stress-repressed genes and that CLAMP inhibits local 3D chromatin loop formation at directly bound loci during heat stress. I identify a specific insulator-remodeler complex comprising Ibf1/2, ZIPIC, and ISWI at direct CLAMP-regulated sites, establishing CLAMP as the first reported DNA-binding transcription factor that directly regulates the repressive role of the heat stress response through modulation of 3D chromatin loops.

In Chapter 4, I leverage publicly available epigenomic data and 3D chromatin interaction data, from the preceding chapters, to build a predictive framework for classifying GA-repeat motifs genome-wide. I constructed a graph in which nodes represent 1kb genomic bins with histone modification features and edges represent significant Micro-C chromatin interactions, then trained a graph con-

volutional neural network (GCN) to distinguish between autosomal, X-chromosomal, and high-affinity site GA-repeats. The GCN substantially outperformed baseline models that lack access to 3D chromatin structure, demonstrating that the spatial context of histone modifications is critical for discriminating between classes of GA-repeat motifs. Feature importance analysis revealed class-specific epigenomic signatures and within-class heterogeneity, providing a framework for identifying the chromatin features that may underlie context-specific pioneer transcription factor occupancy.

In the Discussion, I synthesize the findings from both chapters to propose a unified model in which pioneer TF competition at shared binding motifs drives context-specific co-regulation through differential modulation of 3D chromatin architecture. Together, these chapters provide new mechanistic insight into how differential binding of two essential pioneer transcription factors to similar DNA sequences drives context-specific targeting of transcription complexes and the formation of distinct 3D chromatin environments for coordinate gene regulation.

Chapter 2: Differential chromatin looping regulated by two GA-binding transcription factors creates an X-specific chromatin environment for dosage compensation

Joseph Aguilera^{1,7}, Kaitlyn Cortez^{1,7}, Luna C. Segarra Alonzo², Melissa Aldana^{1,3}, Angelica Aragon¹, Claire Gray^{1,4}, Morgan Woodman-Sousa¹, Kathryn J. Grive^{1,6}, Mukulika Ray^{1,5}, and Erica Larschan¹

Kaitlyn Cortez and I wrote this chapter with editorial feedback from Dr. Erica Larschan.

This chapter is currently under review for publication at *Genome Research*.

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

²Department of Biology, University of Puerto Rico, Río Piedras, San Juan, PR

³Department of Pharmacology and Cancer Biology, Duke University, Durham, NC

⁴Department of Human Genetics, University of Chicago, Chicago, IL

⁵Department of Biology, University of Massachusetts, Boston, Boston, MA

⁶Women and Infants Hospital of Rhode Island, Department of Obstetrics and Gynecology, Program in Women's Oncology, Providence, RI

⁷These authors contributed equally

2.1 Abstract

The mechanisms by which differential occupancy of transcription factors (TFs) at similar binding sites leads to context-specific targeting of large transcription complexes remain poorly understood. X chromosome upregulation (XCU), the most highly conserved step in dosage compensation and best studied in *Drosophila*, serves as a model for understanding how differential occupancy of similar TFs functions context-specifically. Sequence variation within GA-repeat motifs that accumulated on the X chromosome over evolutionary time promotes the binding of a specific GA-binding TF (CLAMP) that recruits the dosage compensation complex (DCC) while outcompeting another similar TF (GAF). However, the mechanism by which CLAMP-GAF competition drives specific targeting of the DCC to the X chromosome remains unknown. Because DCC binding sites cluster in 3D space, we combined Micro-C and Hi-ChIP to determine that CLAMP and GAF directly mediate largely mutually exclusive 3D genomic contacts. Specifically, we show that CLAMP but not GAF drives local short-range interactions that directly link high affinity DCC binding sites with active, dosage-compensated housekeeping genes. In contrast, GAF mediates interactions between transcriptionally silent insulator regions on the X chromosome spanning a wider range of genomic distances. Together, these findings demonstrate that CLAMP outcompetes GAF at active regions on the X chromosome, but not autosomes, to create an X-chromosome specific chromatin environment for dosage compensation. Overall, we provide new insight into how differential TF binding at similar binding sites drives context-specific targeting of transcription complexes.

2.2 Introduction

Within the nucleus of eukaryotic cells, chromatin is tightly compacted through multiple orders of organization (Bonev and Cavalli 2016) and shares finite nuclear space with many classes of transcriptional regulators, including DNA-binding transcription factors (TFs) and large transcription complexes which are often targeted to chromatin by TFs (Zaret 2020). TFs recruit large transcription complexes to facilitate spatially precise and temporally dynamic transcriptional repression and activation of specific groups of genes (Zaret 2020; Drouin 2014; Wu 1995). Furthermore, TFs such as CTCF are important for defining boundaries between three dimensional domains by regulating

the loop extrusion process among other functions (Barth et al. 2025) (Merkenschlager and Nora 2016). In mammals, additional factors that bind to similar binding sites function redundantly with CTCF but their relationship with CTCF remains poorly understood (Skok et al. 2022).

In *Drosophila*, TFs and transcription complexes play an important role facilitating chromatin loops that bring together specific groups of genes for organization and transcription through the function of insulators and tethering elements (Gutierrez-Perez et al. 2019; Boettiger et al. 2016; Melnikova et al. 2004; Wolle et al. 2015; Kaye et al. 2017; Li et al. 2023; Varisco et al. 2025; Levo et al. 2022). Furthermore, recent work indicates that loop extrusion mechanisms function together with tethering elements containing GA-rich sequences to promote loop formation and transcriptional bursting (Choppakatla et al. 2026). While loop extrusion increases the rate of the search for the correct 3D contact, tethering elements are essential to stabilize these contacts and promote transcriptional bursting through heterotypic interactions between different TFs (Varisco et al. 2025; Choppakatla et al. 2026). However, the mechanisms by which specific TFs bring together the correct target loci in a context-specific manner to precisely coregulate genes remain unknown across species.

To understand how specific TFs mediate chromatin loops context-specifically to drive specific transcriptional programs, it is important to use a model in which: 1) specific chromatin loops mediate a particular transcriptional program; 2) a specific TF has been identified that regulates the transcriptional program excluding the function of another known TF. Many TFs that are members of similar functional classes occupy similar binding sites (Zaret 2020). Furthermore, variation within these sites, including SNPs associated with disease, can drive differential occupancy of TFs to similar binding sites (Jeong and Bulyk 2025; Khetan et al. 2025; Jaeger et al. 2010; Kuzu et al. 2016). Yet, in most contexts little is known about the differential functions of these specific TFs in mediating different transcriptional programs or their direct function in regulating 3D loop formation.

Therefore, we used the established *Drosophila* X-chromosome upregulation (XCU) dosage compensation system in which the presence of a specific GA-binding TF has a direct role in mediating the recruitment of the dosage compensation complex (DCC) (Soruco et al. 2013; Larschan et al. 2012; Eggers et al. 2023; Eggers and Becker 2021) compared with another similar factor (Kaye et al. 2018). XCU, which equalizes the levels of X-chromosome and autosomal gene expression within heterogametic species, is the most highly conserved step in dosage compensation across

species that maintains the stoichiometry of the many protein complexes with components encoded on both the X chromosome and autosomes (Samata and Akhtar 2018). Recent work has identified convergent evolution toward the enrichment of the H4K16ac chromatin mark which increases transcript levels two-fold as the most highly conserved mechanism for dosage compensation which occurs across heterogametic species (Zimmer et al. 2025; Gu et al. 2019). Because XCU was first identified in *Drosophila*, which has excellent tools for dissecting transcriptional mechanisms, it has been best studied in this system (Samata and Akhtar 2018).

Variation within GA-rich binding sites that accumulated over evolutionary time on the X chromosome promotes the occupancy of the pioneer TF CLAMP (Chromatin Linked Adaptor for MSL Proteins) instead of the well-studied pioneer TF GAGA-Factor (GAF) (Kaye et al. 2018; Kuzu et al. 2016). CLAMP mediates DCC binding to the X chromosome to facilitate the two-fold upregulation of the single male X chromosome (Soruco et al. 2013; Eggers et al. 2023). Apart from dosage compensation, these two TFs have opposing transcriptional functions at most genomic loci: GAF often acts as a transcriptional activator (Adkins et al. 2006; Duarte et al. 2016), while CLAMP often functions as a repressor (Urban et al. 2017a). Therefore, understanding the competitive relationship between GAF and CLAMP will also reveal new insight into how the balance between activating and repressive transcriptional programs is established across the genome.

Despite decades of study, the mechanisms by which an X-chromosome-specific chromatin environment is established to recruit the DCC to the bodies of active genes remains poorly understood (Larschan et al. 2007; Alekseyenko et al. 2012). CLAMP binds to GA-rich sequence motifs on the X chromosome within the strongest DCC binding sites which are known as chromatin entry sites (CES) (Kelley et al. 1999) or high-affinity sites (HAS) (Straub et al. 2008). Subsequent to identifying these binding sites that are already occupied by CLAMP in early development (Rieder et al. 2019), the DCC spreads along the length of the X chromosome through the function of DCC components, including the roX non-coding RNAs (Rieder et al. 2019; Valsecchi et al. 2020; Park et al. 2002; Kelley et al. 1999).

Beyond its function in XCU, CLAMP is a pioneer TF that binds to GA-rich motifs to regulate chromatin accessibility and remodeling in other contexts such as zygotic genome activation (Duan et al. 2021) and histone gene transcription (Rieder et al. 2017). GAF is one of the first TFs to be identified across species and recognizes GA-rich motifs throughout the genome to mediate

activation of transcription in many contexts including zygotic genome activation (Gaskill et al. 2021) and regulation of the heat shock genes (Duarte et al. 2016; Adkins et al. 2006). Variation in sequence composition at GA-motifs determines whether CLAMP or GAF will occupy a specific site, with CLAMP but not GAF tolerating variation with stretches of GA-repeats that occurs within the strongest DCC binding sites (HAS/CES) (Kaye et al. 2018; Adkins et al. 2006). While CLAMP but not GAF directly associates with the DCC both *in vivo* and *in vitro* (Alekseyenko et al. 2015; Eggers et al. 2023), GAF promotes DCC recruitment but to a much lesser extent (Kaye et al. 2018). Therefore, XCU serves as a model for understanding how small differences in DNA sequence determine the context-specific recruitment of large transcription complexes that induce coordinated regulation of specific groups of genes.

Prior low-resolution studies suggested that the strongest DCC binding sites cluster in 3D space and CLAMP contributes to this process (Jordan and Larschan 2021). GAF has also been shown to contribute to 3D chromatin looping in the nervous system (Li et al. 2023). Therefore, we hypothesized that the competition between CLAMP and GAF leads to context-specific formation of chromatin looping that differs on the X chromosome compared to autosomes. To define how factors with similar binding sites differentially regulate chromatin loop formation, we determined the individual direct functions of GAF and CLAMP on the X chromosome and autosomes by conducting both Micro-C and Hi-ChIP. We first identified CLAMP and GAF binding sites using Hi-ChIP, then identified differential chromatin looping within the nucleus after knocking down GAF and CLAMP individually using Micro-C. By integrating Hi-ChIP and Micro-C we define direct roles that CLAMP and GAF play in chromatin loop formation at up to 1kb resolution.

Specifically, we demonstrate that the differential occupancy of CLAMP versus GAF which is caused by minor sequence variation within GA-rich motifs on the X chromosome (Kaye et al. 2018) drives the formation of local short-range interactions that can directly link high affinity DCC binding sites with active housekeeping genes that are dosage compensated. In contrast, GA-repeat motifs that are contiguous promote GAF binding that can drive interactions between silent insulator regions on the X chromosome and span a broad range of genomic distances. On autosomes, both CLAMP and GAF directly regulate contacts between active chromatin regions including active promoters. Therefore, CLAMP outcompetes GAF at active regions on the X chromosome but not autosomes to mediate the short-range 3D contacts required for DCC recruitment specifically

to active genes on the X-chromosome that need to be dosage compensated. CLAMP but not GAF can directly interact with the DCC to recruit it to its target genes where it upregulates transcription (Eggers et al. 2023; Eggers and Becker 2021), suggesting that these 3D contacts are important for DCC occupancy (Sorucu et al. 2013) and transcriptional upregulation (Urban et al. 2017a) that are both mediated by CLAMP. Overall, we provide key insight into how differential binding of two essential TFs to similar binding sites drives context-specific targeting of a transcription complex.

2.3 Results

2.3.1 CLAMP- and GAF-dependent chromatin loops are largely non-overlapping and have different size distributions

To compare how CLAMP and GAF regulate chromatin loop formation, we used *Drosophila* S2 cells as a model system for the following reasons. S2 cells are embryonically derived male *Drosophila* cells that have been used to study XCU for several decades (Straub et al. 2008; Alekseyenko et al. 2012, 2008; Valsecchi et al. 2020). The DCC binds to very similar loci in S2 cells compared with embryos and neuronally derived cell lines. This suggests that using S2 cells provides sufficient biological context while also circumventing asynchronous embryonic lethality that occurs after the loss of maternally derived CLAMP in embryos (Rieder et al. 2017; Alekseyenko et al. 2008, 2012).

To define the function of the CLAMP and GAF GA-binding transcription factors in shaping the 3D chromatin architecture of the *Drosophila* genome, we utilized a high-resolution chromatin mapping technique, Micro-C (Hsieh et al. 2015), with established RNAi-mediated knockdown approaches (Sorucu et al. 2013; Larschan et al. 2012; Kaye et al. 2018) in *Drosophila* S2 cells (Fig. 2.1A). We validated *clamp* and *gaf* RNAi knockdowns using Western blotting (Fig. S2.1). Whole-chromosome differential contact maps revealed that depletion of either CLAMP or GAF disrupts genome-wide looping, substantially remodeling the 3D architecture of the *Drosophila* genome (Fig. S2.2).

The depletion of the GA-binding transcription factors CLAMP and GAF profoundly disrupted higher-order chromatin organization. We first identified 3,852 TADs in control conditions (Fig. S2.3A). Depletion of CLAMP altered 1,218 (31.62%) of these TADs, while GAF depletion had an even more dramatic effect, disrupting 3,415 (88.66%) of all TADs (Fig. S2.3A, C). Notably, 1,044

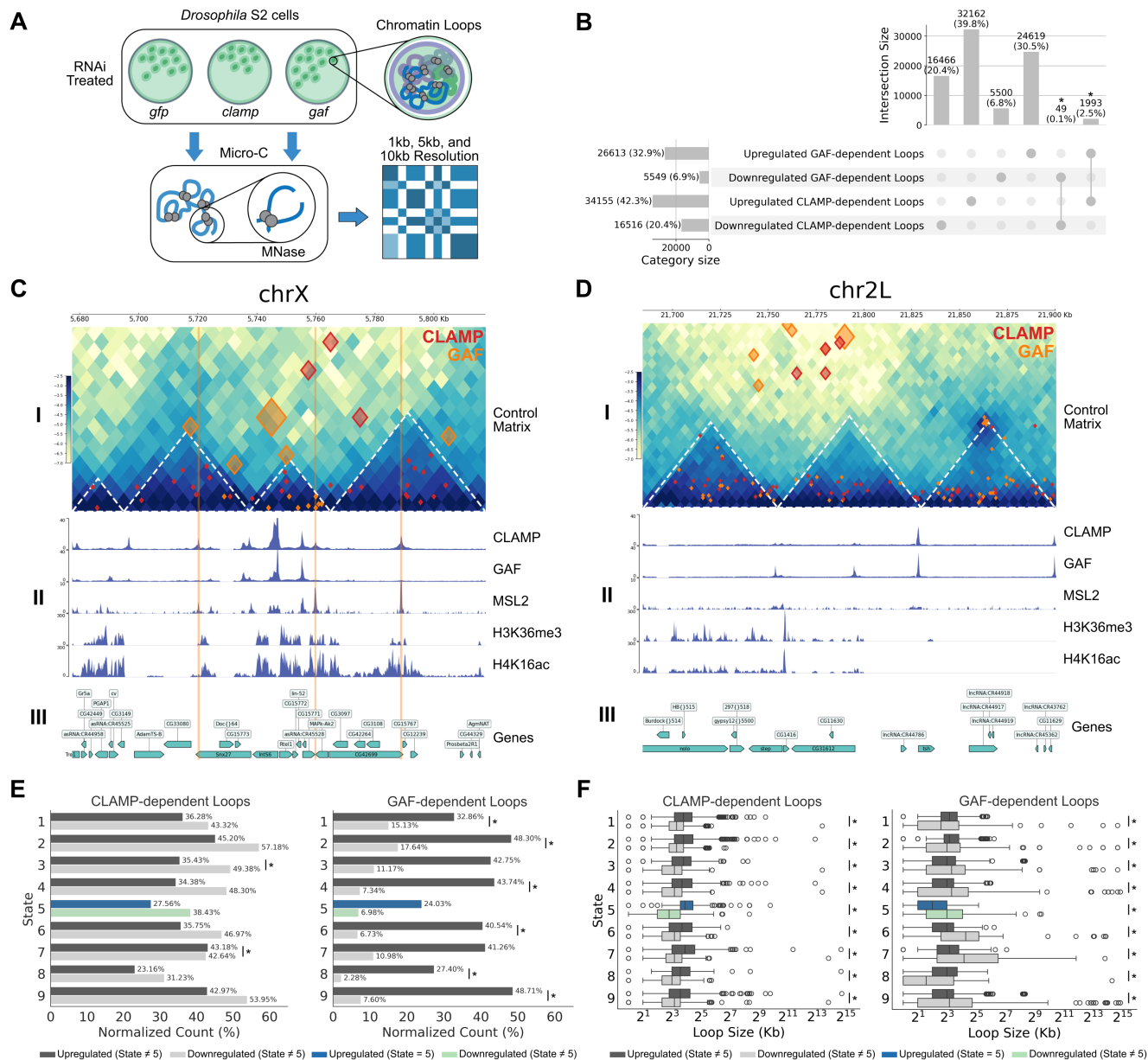


Figure 2: Loss of CLAMP versus GAF alters different chromatin loops. (A) Experimental workflow for Micro-C analysis. Schematic illustrating RNAi-mediated depletion of CLAMP or GAF in *Drosophila* S2 cells, with *gfp* RNAi serving as a control. Micro-C libraries were generated using MNase digestion and sequenced at 1kb, 5kb, and 10kb resolutions. (B) Loops altered by loss of CLAMP or GAF overlap at a low frequency. UpSet plot showing the intersection of significantly altered chromatin loops (Fisher's Exact Test, $P < 0.05$) following *clamp* RNAi (*clampi*) or *gaf* RNAi (*gafii*) treatment. Vertical bars represent the number of loops that are either upregulated (gained) or downregulated (lost) across the indicated categories. (C-D) Representative Micro-C contact maps and genomic tracks. CoolBox plots displaying the chromatin landscape on (C) the X chromosome (chrX) and (D) an autosome (chr2L). (I) Control Micro-C contact matrices (*gfp*) showing differential CLAMP- (red diamonds) and GAF-dependent (orange diamonds) loops. Size of diamond corresponds to the resolution of the differential chromatin interaction: small (1kb), medium (5kb), and large (10kb). (II) Normalized ChIP-seq tracks for CLAMP, GAF, MSL2, H3K36me3, and H4K16ac across a representative 200 kb genomic window. (III) *Drosophila* (*dm6*) genes are shown at the bottom; boxes represent exons and arrows indicate transcriptional orientation. (E) Chromatin state enrichment of differential loops. Normalized counts (%) of CLAMP- and GAF-dependent loops categorized by the 9-state chromatin model. Asterisks (*) indicate significant difference in enrichment (Fisher's Exact Test, $P < 0.05$). (F) Genomic length span of state-specific loops. Boxplots showing the loop size (kb) distribution for CLAMP- and GAF-dependent loops across chromatin states 1-9. Asterisks (*) indicate significant difference between loop sizes (Mann-Whitney Test, $P < 0.05$). For panels E and F, upregulated loops in State 5 (dosage compensation) are highlighted in blue, while downregulated State 5 (dosage compensation) loops are highlighted in green.

of these domains, representing 27.1% of all control TADs, were disrupted in both knockdowns, a highly significant overlap ($P < 0.001$, Fisher's Exact Test) (Fig. S2.3B). The widespread loss of TAD integrity after CLAMP and GAF depletion underscores the critical role of these TFs in maintaining chromatin architecture.

To maximize the high-resolution capabilities of Micro-C, we performed differential chromatin loop analyses at three resolutions (1kb, 5kb, and 10kb) for each knockdown condition (*clamp* or *gaf* RNAi) to detect both short- and long-range differential loops (Fig. 2.1A). We integrated the multi-resolution loops for each condition which we then used throughout our analysis (Table S2.1).

Differential loop quantification with the HiCcompare R-package (Stansfield et al. 2018) revealed four distinct sets of loops (Fig. 2.1B): 1) chromatin loops that are upregulated after *clamp* RNAi relative to control *gfp* RNAi (upregulated CLAMP-dependent loops, N=34,155); 2) chromatin loops that are downregulated after *clamp* RNAi relative to control *gfp* RNAi (downregulated CLAMP-dependent loops, N=16,516); 3) chromatin loops that are upregulated after *gaf* RNAi relative to control *gfp* RNAi (upregulated GAF-dependent loops, N=26,613); 4) chromatin loops that are downregulated after control *gaf* RNAi relative to control *gfp* RNAi (downregulated GAF-dependent loops, N=5,549). Across the four sets of differential chromatin loops, we only observed very minimal direct overlap between the following groups (Fig. 2.1B): upregulated CLAMP- and GAF-dependent loops (49 loops; 0.1%; $P < 0.01$, Fisher's Exact Test) and downregulated CLAMP- and GAF-dependent loops (1,993 loops, 2.5%, $P < 0.01$, Fisher's Exact Test). The minimal direct overlap observed between differential CLAMP- and GAF-dependent loops suggests that CLAMP and GAF have largely mutually exclusive looping roles in the genome even though they sometimes regulate loops within the same TADs.

Interestingly, CLAMP depletion resulted in far fewer differential TADs than GAF despite yielding substantially more differential loops (Fig. 2.1B, Fig. S2.3). Furthermore, 85.71% of TADs disrupted by CLAMP knockdown were also disrupted by GAF knockdown, whereas only 30.57% of GAF-disrupted TADs overlapped with those affected by CLAMP depletion (Fig. S2.3). These data suggest that GAF may have a more prominent role in maintaining TAD integrity than CLAMP. The observation that CLAMP depletion produces more differential loops, yet fewer disrupted TADs, implies that CLAMP may primarily regulate precise chromatin looping within TADs, whereas GAF is more critical for establishing or maintaining the broader domain boundaries that define TAD

structure.

Subsequently, we performed a comparative visualization of a 220.5kb X-chromosome region (chrX:5,604,766-5,825,266) and an 185kb autosomal region (chr2L:21,660,000-21,845,000). For each region, we examined differential chromatin loops relative to local TADs, CLAMP and GAF occupancy, MSL2 (the DCC core protein subunit), histone modifications, and the *Drosophila* S2 9-state chromatin model (Fig. 2.1C, D).

A key looping characteristic emerged through visualization in both genomic contexts, revealing a resolution-dependent distinction between intra-TAD and inter-TAD loop organization. On the X chromosome, we first visualized all differential loops over the control (*gfp* RNAi) Micro-C contact matrix and its corresponding TADs, with the region's High Affinity Sites (HAS) highlighted in orange to contextualize the 3D chromatin changes and factor occupancy (Fig. 2.1C: I). With this visualization established, we observed that high-resolution (1kb) differential loops dependent on CLAMP and GAF were primarily contained within TADs, while lower-resolution (5kb and 10kb) loops often connected regions outside of these domains. Notably, this same resolution-dependent characteristic was also present on the autosomal region, where 1kb loops were intra-TAD and lower-resolution loops extended to inter-TAD interactions (Fig. 2.1D: I). This consistent resolution-dependent partitioning of intra- and inter-TAD loops underscores that multi-resolution analysis is essential to capture the complete and diverse set of architectural changes mediated by these factors, regardless of the chromosome.

When comparing CLAMP and GAF occupancy at the X-chromosome region, our findings are consistent with prior findings that CLAMP outcompetes GAF at HAS which has been demonstrated *in vivo* (Kaye et al. 2018) and *in vitro* (Eggers and Becker 2021; Eggers et al. 2023). CLAMP is enriched while GAF is absent in the highlighted HAS (Fig. 2.1C: II). Importantly, MSL2 is highly enriched at the HAS, consistent with findings from previous studies which outline the synergistic relationship between CLAMP and MSL2 at the HAS (Eggers et al. 2023; Soruco et al. 2013). To mark DCC activity, we observed the presence of H4K16ac (Gu et al. 1998) and visualized the H3K36me3 chromatin mark to show the presence of active gene bodies (Larschan et al. 2007) in this region (Fig. 2.1C: II, III). Importantly, these histone modifications are enriched flanking the HAS, consistent with local spreading of the DCC from its initialization sites (Kelley et al. 1999) visualized in two dimensions. Conversely, the autosomal region lacks both HAS and MSL2 enrichment

as expected (Fig. 2.1D: II, III). While CLAMP and GAF occupy the GA-repeats present throughout the genome, the lack of MSL2 and H4K16ac occupancy on the autosome region underscores the X-specificity of DCC recruitment.

After visualizing differential CLAMP- and GAF-dependent loops over specific candidate regions, we measured how these loops are distributed across different chromatin states globally using the *Drosophila* S2 9-state HMM chromatin model, which bins the genome into one of nine chromatin states by integrating dozens of modENCODE chromatin profiling data sets (Kharchenko et al. 2010). Importantly, we exclusively used differential loops detected at a 1kb resolution for the chromatin state enrichment analysis, as lower resolutions contained too much chromatin state diversity within loop anchors. For each set of differential CLAMP- and GAF-dependent loops, we determined whether there was a difference in chromatin state enrichment between upregulated and downregulated loops. Interestingly, differential CLAMP-dependent loops yielded only two significant enrichment differences, in the open chromatin (state 4) and heterochromatin (state 7), while differential GAF-dependent loops yielded significant enrichment differences across a broader range of chromatin environments, spanning active transcription (states 1 and 2), open chromatin (state 4), Polycomb-mediated repression (state 6), heterochromatin-like regions (state 8) and transcriptionally silent euchromatin (state 9) (Fig. 2.1E). The stark difference in chromatin state enrichment between CLAMP- and GAF-dependent loops, with only the open chromatin state (state 4) shared between them, suggests that these similar GA-binding TFs may regulate functionally distinct chromatin environments across the genome.

After measuring chromatin state enrichment, we asked whether the sizes of differential CLAMP- and GAF-dependent loops differ across chromatin states (Fig. 2.1F). Notably, within every chromatin state category, the size distributions of upregulated and downregulated loops were significantly different for both CLAMP and GAF knockdowns (Fig. 2.1F; $P < 0.001$, Mann-Whitney Test). However, in addition to this shared significance, a distinct TF-specific shift in the interquartile ranges (IQRs) emerged. When examining differential CLAMP-dependent loops, the IQRs of upregulated loops expanded and shifted upward relative to downregulated loops, indicating a loss of shorter-range interactions upon CLAMP depletion and a gain of longer-range interactions. In contrast, the opposite pattern was observed for differential GAF-dependent loops: the IQRs of the upregulated loops contracted and shifted downward relative to downregulated loops, indicating

a loss of longer-range interactions upon GAF depletion and a gain of shorter-range interactions. Together, these opposing size shifts suggest that CLAMP and GAF have different effects on chromatin looping: CLAMP loss preferentially disrupts local interactions, whereas GAF loss disrupts longer range interactions.

Importantly, the global looping characteristics described above for CLAMP- and GAF-dependent loops also occurred at dosage compensated genes (state 5) (Fig. 2.1E, F). Given that *Drosophila* dosage compensation is regulated by well-characterized DNA regulatory elements (Alekseyenko et al. 2008, 2012), TF competition (Kaye et al. 2018), and a large chromatin modifying complex (Samata and Akhtar 2018), we can leverage this conserved biological process (Zimmer et al. 2025; Deng et al. 2013) to understand how GAF and CLAMP function differentially to regulate transcription context-specifically.

2.3.2 CLAMP and GAF are required for different sized chromatin loops near GA-repeat rich high-affinity sites (HAS)

Due to the established competition between CLAMP and GAF high affinity sites (HAS) that regulates DCC recruitment (Kaye et al. 2018), we next examined how CLAMP and GAF regulate chromatin looping specifically at HAS. We quantified the proportion of dosage-compensated genes contained within the differential CLAMP- and GAF-dependent loops, and which were also associated with the HAS. We determined that CLAMP-dependent loops encompass over 90% of dosage-compensated (DC) genes anchored at HAS on the X chromosome, whereas GAF-dependent loops only encompassed 50% or fewer DC genes associated with HAS (Fig. 2.2A). Furthermore, CLAMP-dependent loops were found to overlap the HAS more frequently than GAF-dependent loops (Fig. 2.2A). These data are consistent with the established stronger requirement for CLAMP in regulating DCC recruitment to HAS compared with GAF which also has a significant although smaller role which had not yet been defined (Kaye et al. 2018).

Therefore, we dissected the differences between GAF and CLAMP regulated loops at HAS because the loops they regulate encompass different percentages of DC genes. Micro-C APA plots revealed robust looping at the HAS (log enrichment = 1.74) that diminished substantially upon CLAMP knockdown (log enrichment = 1.58), while GAF depletion caused a more diffuse loss of signal distal at the HAS (log enrichment = 1.70) (Fig. 2.2B).

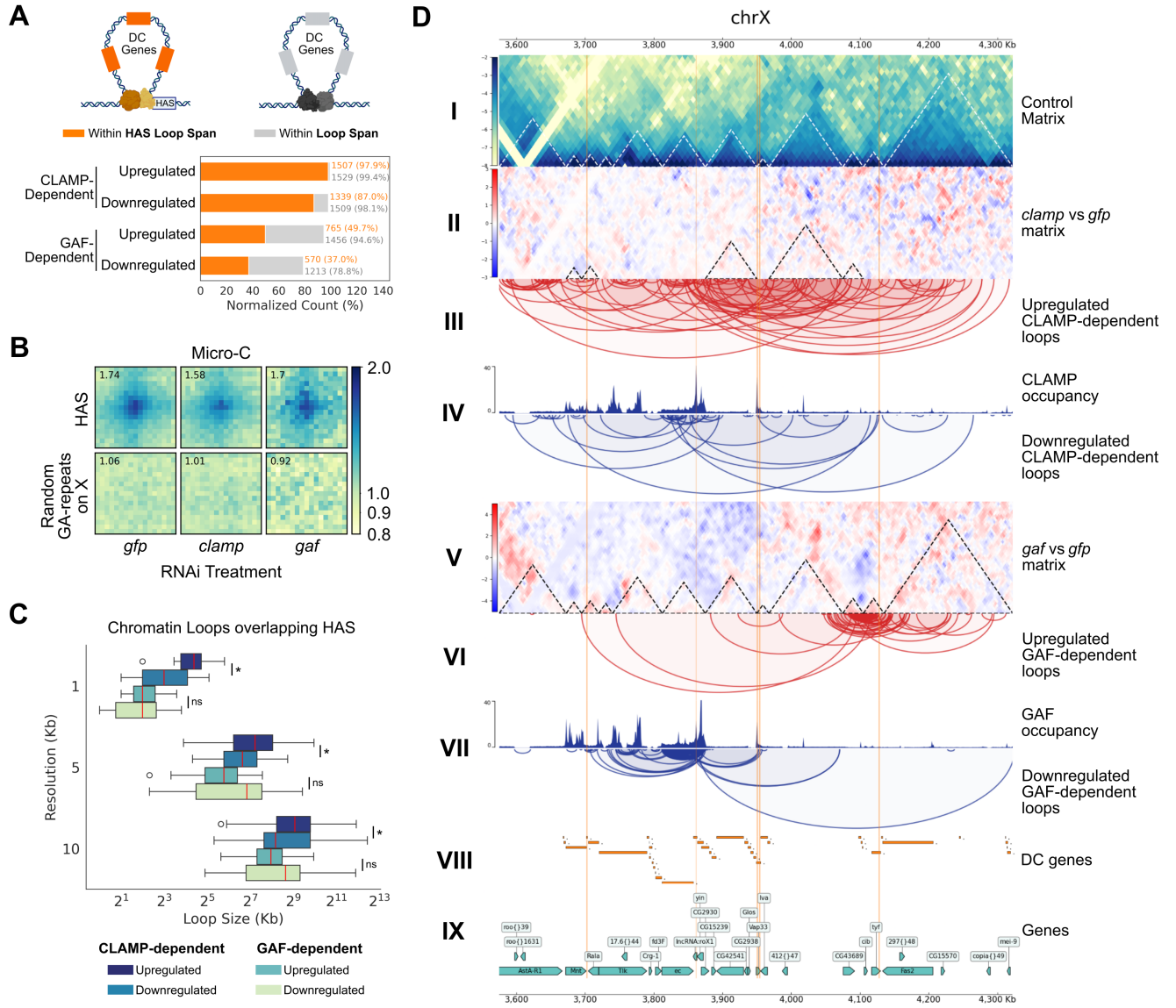


Figure 2.2: Loss of CLAMP and GAF alters chromatin loop architecture at High Affinity Sites (HAS) on the X chromosome. (A) Association of CLAMP- and GAF-dependent loops with dosage-compensated (DC) genes. Top: Schematic of: Left: DC genes (orange) localized within the span of loops anchored at High Affinity Sites (HAS). Right: DC genes not within HAS loop span. Bottom: Normalized counts (%) of upregulated and downregulated CLAMP- and GAF-dependent loops containing DC genes within their span. Numbers indicate the total DC gene count and the percentage of DC genes that are encompassed by loops either associated with HAS or not. (B) Aggregate Peak Analysis (APA) of HAS contacts. Heatmaps showing the average Micro-C contact enrichment centered on High Affinity Sites (top row) compared to random GA-repeat sequences on the X chromosome (bottom row). Enrichment scores (upper left) indicate a reduction in contact strength following *clamp* RNAi (enrichment = 1.58) or *gaf* RNAi (enrichment = 1.7) treatment compared to *gfp* RNAi control (enrichment = 1.74). Pixel resolution is 2.5 kb with ± 20 kb flanking central pixel. (C) Distribution of loop sizes at HAS. Box plots categorized by Micro-C resolution (1kb, 5kb, 10kb) showing the genomic span (Kb) of upregulated and downregulated CLAMP- and GAF-dependent loops overlapping HAS. (D) Representative chromatin loops regulated by CLAMP or GAF on the X chromosome at HAS (CoolBox visualization) (I) Control Micro-C contact matrix (*gfp*). (II & V) Differential contact matrices showing log2 fold-changes for *clampi* vs *gfpi* and *gafi* vs *gfpi*. Red indicates increased contacts and blue indicates decreased contacts. Differential TADs are shown with black dashed lines. (III, IV, VI, VII) Loop tracks representing upregulated (red) and downregulated (blue) loops for CLAMP and GAF, respectively, aligned with their corresponding protein occupancy Hi-ChIP tracks. (VIII) Genomic tracks for DC genes and (IX) dm6 gene annotations.

Additionally, we observed a distinct relationship between loop sizes: After CLAMP knockdown, downregulated loops over HAS were considerably smaller than upregulated loops (Fig. 2.2C; $P < 0.05$, Mann-Whitney Test). The 1kb resolution provided by Micro-C reveals CLAMP anchors precisely at the center of HAS at the nucleosomal-scale level (Fig. 2.2C; $P < 0.001$; Mann-Whitney Test) that may have otherwise been blurred at 10kb or masked by multi-resolution aggregation (Fig. 2.2C; $P < 0.05$; Mann-Whitney Test). Conversely, GAF knockdown loop distribution over HAS showed that downregulated loops maintained a larger median loop size than upregulated loops (Fig. 2.2C). This variation in loop size was best observed at 5kb and 10kb resolutions, where larger loops can be captured (Fig. 2.2C). These findings suggest that CLAMP and GAF may maintain distinct, size-specific regulatory roles at the HAS, an observation strongly correlated with the specific loop size ranges identified by Micro-C.

Examination of a 747 kb candidate region on the X chromosome (chrX:3,575,000-4,322,000) further illustrated the divergent effects of CLAMP and GAF loss on chromatin looping near HAS (Fig. 2.2D). The increase in contacts surrounding and within differential TADs upon CLAMP knockdown demonstrates that CLAMP is necessary for maintaining short, high-frequency loops at HAS (Fig. 2.2D: I, II, IV). Conversely, the loss of larger chromatin loops around differential TADs upon GAF knockdown demonstrates that GAF regulates the broader neighborhood surrounding HAS rather than only the specific HAS locus (Fig. 2.2D: I, V, VII). Notably, CLAMP depletion resulted in approximately twice as many upregulated than downregulated chromatin loops, a pattern visible within the candidate region (Fig. 2.2D: III). This asymmetry was even more pronounced upon GAF depletion, where upregulated loops outnumbered downregulated loops by approximately five-fold (Fig. 2.2D: VI). Together, our data support a model in which CLAMP and GAF function at HAS to cluster DC genes into 3D domains of different sizes (Fig. 2.2D: VIII, IX). However, it is critical to determine which loops are directly versus indirectly regulated by CLAMP and GAF to define their mechanism of action.

2.3.3 Single nucleotide variation within GA-rich motifs correlates with TF-specific chromatin looping at HAS

We observed that CLAMP and GAF influence loops of different sizes at HAS and throughout the genome (Fig. 2.1F, Fig. 2.2C). However, with Micro-C alone, we were not able to determine

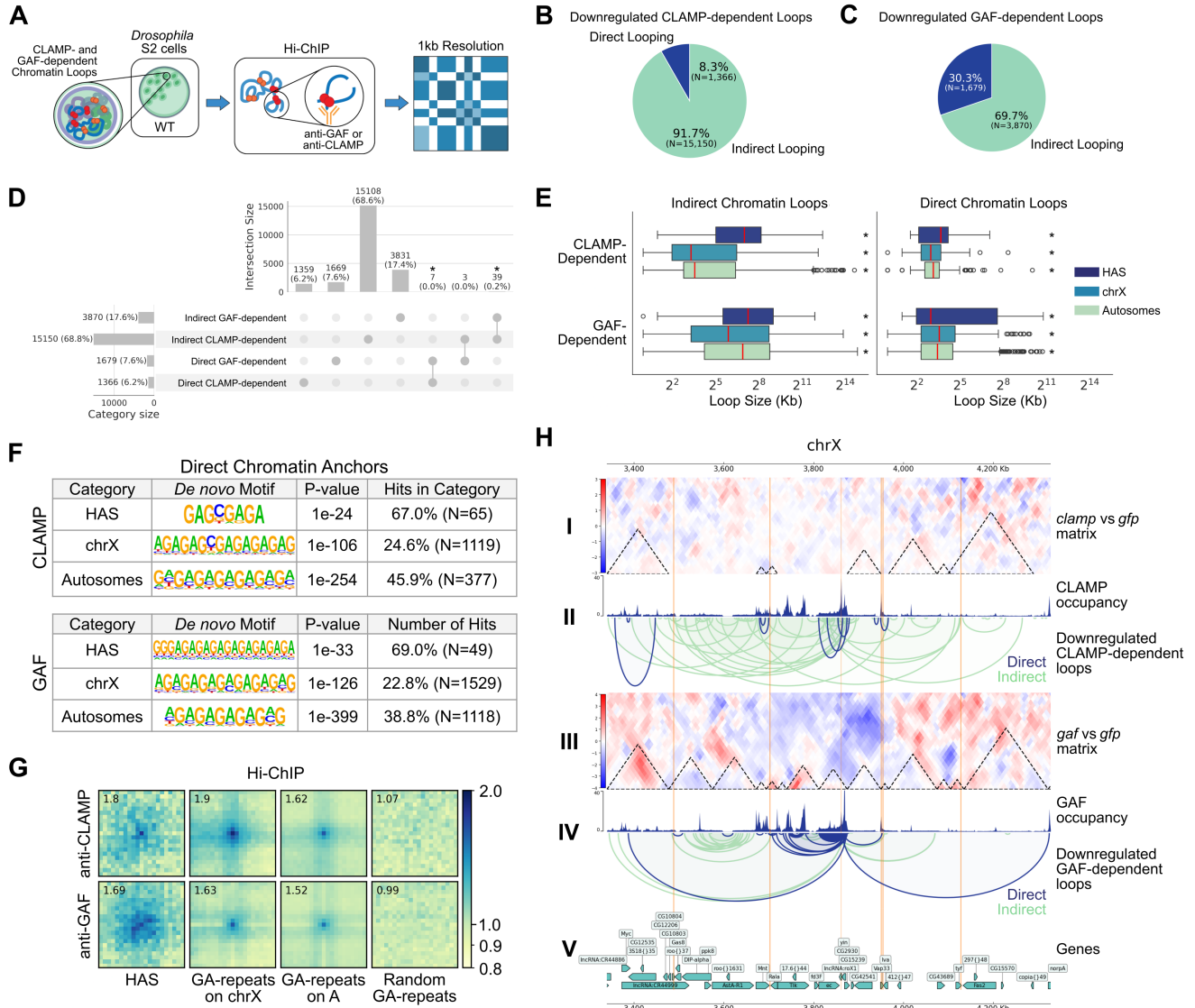


Figure 2.3: Direct looping by CLAMP and GAF suggests size-specific loop regulation at the HAS. (A) Hi-ChIP experimental workflow. Schematic illustrating the Hi-ChIP protocol in *Drosophila* S2 cells using anti-CLAMP and anti-GAF antibodies to identify protein-specific chromatin interactions at a 1kb resolution. (B–C) Frequency of indirect looping versus direct looping. Pie charts showing the proportion of downregulated (B) CLAMP-dependent and (C) GAF-dependent loops that are a direct or indirect effect of their respective protein knockdown. (D) Overlap of direct and indirect CLAMP- and GAF-dependent loops. UpSet plot showing the intersection between directly and indirectly mediated loop categories for both CLAMP and GAF. (E) Direct loop size distribution shows size specificity over HAS. Boxplots comparing the genomic span (kb) of indirect (left) and direct (right) loops across High Affinity Sites (HAS), the X chromosome (chrX), and autosomes. Direct loops at HAS exhibit a more constrained size distribution compared to those on autosomes. Asterisks (*) indicate significant difference between direct and indirect loop sets (Mann-Whitney Test, $P < 0.05$). (F) *De novo* motif enrichment at direct anchors. Tables listing the top *de novo* motifs at direct CLAMP and GAF loop anchors. (G) Hi-ChIP Aggregate Peak Analysis (APA). APA plots showing enrichment of CLAMP and GAF binding at HAS and GA-repeats on the X chromosomes and autosomes relative to random GA-repeats (control). Pixel resolution is 1.0kb with ± 10 kb flanking central pixel. (H) Visualization of direct vs indirect looping on the X chromosome. CoolBox plots displaying: (I & III) Differential contact matrices (log2FC) showing changes in TAD architecture following *clamp* and *gaf* RNAi treatment. (II & IV) Normalized Hi-ChIP occupancy tracks coupled with annotated direct (blue) and indirect (green) chromatin loops. (V) dm6 gene annotations on the X chromosome.

which differential loops were a result of a direct absence of either TF or an indirect effect mediated by other regulators. Therefore, we incorporated the MNase Hi-ChIP technique (Mumbach et al. 2016) targeting CLAMP and GAF to annotate the Micro-C differential loops to distinguish direct from indirect effects of TF-specific knockdowns (Fig. 2.3A). Notably, direct and indirect effects were only determined for the downregulated CLAMP- and GAF-dependent loops as these are the interactions that are present under control conditions which are the same conditions we used for Hi-ChIP.

By integrating Micro-C with Hi-ChIP, we were able to assign with high confidence which down-regulated CLAMP- and GAF-dependent loops were directly or indirectly regulated by each factor. Furthermore, we defined direct looping as either factor being present at one or both of the two loop anchors and indirect looping as neither factor being present at either loop anchor. We characterized the following four groups of loops (Fig. 2.3B, C): 1) direct downregulated CLAMP-dependent loops (N=1,366, 8.3%), 2) indirect downregulated CLAMP-dependent loops (N=15,150, 91.7%), 3) direct downregulated GAF-dependent loops (N=1,679, 30.3%), and 4) indirect downregulated GAF-dependent loops (N=3,870, 69.7%). As we had expected, there are fewer direct than indirect loops consistent with CLAMP and GAF being essential proteins which regulate the expression and function of many other regulators in a cellular context (Urban et al. 2017b; Adkins et al. 2006).

Next, we wanted to know whether there was overlap between the direct and indirect loop sets. We performed an intersection of all loops and determined that, although minimal, there was significant overlap between direct CLAMP- and GAF-dependent loops (N=7, 0.23%) and indirect CLAMP- and GAF-dependent loops (N=39, 0.21%) (Fig. 2.3D; $P < 0.001$, Fisher's Exact Test). This indicates that there are a few loci which CLAMP and GAF directly or indirectly influence together. However, given that the magnitude of the overlap is very small, we suggest that these two pioneer TFs possess distinct genome-wide looping roles although they bind to very similar binding sites (Fig. 2.3D).

We next asked whether loops directly mediated by CLAMP or GAF have a differential size distribution compared with indirect loops and each other. Interestingly, the distribution of loop sizes for the indirect CLAMP- and GAF-dependent loops are all significantly larger and contain wider interquartile ranges in comparison to the corresponding direct loop categories (Fig. 2.3E; $P < 0.001$, Mann-Whitney Test). Interestingly, we observed that unlike other loop classes direct GAF-

dependent loops that overlap HAS have a large and wide interquartile range, similar to the indirect loop set (Fig. 2.3E, $P < 0.001$, Mann-Whitney Test). This indicates that GAF is directly responsible for a wide range of short- to long-range interactions at the HAS, while promoting more uniform short- to medium-sized interactions at other sites on the X chromosome and autosomes. In contrast, the direct CLAMP-dependent loops exhibit a narrow interquartile range and are short in size, suggesting the direct role of CLAMP is mediating short-range local chromatin loops at the HAS, X chromosome, and autosomes. Lastly, the direct GAF-dependent loops are significantly larger than the direct CLAMP-dependent loops on the X chromosome and autosomes (Fig. 2.3E; $P < 0.001$, Mann-Whitney Test). Notably, the direct GAF-dependent loops exhibited a larger loop-size range at the HAS than direct CLAMP-dependent loops (Fig. 2.3E), but the total number of chromatin loops exhibiting this difference was not enough to provide statistical power to the comparison of the medians. Together, these data suggest that CLAMP regulates the formation of a shorter range and largely mutually exclusive set of chromatin loops compared with GAF.

To identify DNA motifs present at directly regulated loop anchors, we intersected CLAMP and GAF ChIP peaks with their respective loop anchors, thereby retaining only TF-bound loop anchors. Subsequently, we utilized the HOMER motif discovery and analysis suite (Heinz et al. 2010) to identify *de novo* motifs. We report the top *de novo* motif found in each site category using the direct CLAMP- and GAF-dependent TF-bound anchors (Fig. 2.3F). The most striking difference in motifs is between CLAMP- and GAF-dependent anchors at the HAS, where GA-repeats are interrupted by a single highly conserved cytosine nucleotide (67.3% frequency in Motif PWM) on CLAMP-bound anchors but not GAF-bound anchors (Fig. 2.3F, upper). This single cytosine interruption is also maintained in the top motif detected in CLAMP-dependent anchors on other sites on the X chromosome (61.6% frequency in Motif PWM). In the remaining site categories, the break within the GA-repeat motif is less pronounced as we observed cytosine conservation increase only modestly at other locations in the genome. These results are consistent with our prior findings that a cytosine residue within the GA-repeat favors CLAMP compared with GAF binding (Kaye et al. 2018). Together, these results suggest that CLAMP and GAF bind GA-repeats that differ by a single nucleotide where they drive the formation of loops of different sizes.

Next, we asked whether the pattern of chromatin looping at GA-repeat anchors differs between direct CLAMP- and GAF-dependent loops at HAS and other GA-repeats throughout the genome.

Given that GA-repeats are abundant in the genome and have multiple functions, we restricted our analysis to GA-repeats located at the loop anchors of CLAMP or GAF direct loops (either at HAS or other genomic locations), thereby isolating GA-regulatory elements engaged in TF-mediated chromatin looping. Leveraging CLAMP- and GAF-specific Hi-ChIP contact matrices, we used APA plots to measure TF-specific aggregated looping over three classes of sites: HAS, GA-repeats on the X chromosome, and GA-repeats on the autosomes (Fig. 2.3G). Subsequently, we created union sets of the GA-repeats which interact with CLAMP or GAF since either TF may interact with the HAS.

We observed that CLAMP has a strong enrichment of direct loops at the HAS, with the central pixel (1kb bin) showing substantial contact frequency relative to flanking regions (Fig. 2.3G). This looping signal is consistent with prior work demonstrating that CLAMP binds directly to the HAS (Soruco et al. 2013; Eggers and Becker 2021), and extends these findings by revealing that CLAMP not only recognizes the HAS but also mediates chromatin looping at these sites. In contrast, GAF showed a broader, less focused looping pattern at the HAS, with enrichment dispersed across a ± 3 kb window rather than concentrated at the central pixel (Fig. 2.3G). This diffuse signal, while still centered on the HAS, is consistent with prior reports that GAF contributes to DCC binding but it is not nearly as strongly required as CLAMP (Soruco et al. 2013) (Kaye et al. 2018). Together, these results suggest that the ability of CLAMP to mediate precise, focal chromatin loops at the HAS may be critical for efficient DCC recruitment, whereas GAF may stabilize or reinforce these interactions over a broader spatial range.

Next, we examined direct looping by CLAMP and GAF at GA-repeats on the X chromosome and autosomes (outside of HAS) to define context-specific functions of these factors. CLAMP preferentially regulated looping at X-linked GA-repeats compared to autosomal GA-repeats, consistent with its known enrichment on the male X chromosome (Soruco et al. 2013; Kaye et al. 2018). In contrast, GAF exhibited similar looping strength at GA-repeats on both the X chromosome and autosomes, with only a modest increase on the X, suggesting that GAF functions more uniformly across chromosomes at these elements. Interestingly, aggregated loops directly regulated by CLAMP displayed a striped signal pattern at GA-repeats on both the X chromosome and autosomes (Fig. 2.3G), indicating that CLAMP promotes broad, directional chromatin interactions stemming from GA-rich loci rather than forming discrete point-to-point loops (Vian et al. 2018). In

contrast, loops directly regulated by GAF showed no detectable striping signal at GA-repeats on either chromosome, suggesting that GAF organizes more focal chromatin contacts without the extended directional interactions that CLAMP exhibits. Together, our results suggest that CLAMP and GAF have distinct context-specific direct functions in mediating looping at HAS, X-chromosome, and autosomal GA-repeats.

To visualize the differential characteristics of direct versus indirect CLAMP- and GAF-mediated looping, we examined a 988 kb region of the X chromosome (chrX:3,340,000-4,328,000) containing HAS, TF-specific loops, and differential TADs (Fig. 2.3H). CLAMP-dependent loops in this region recapitulated global patterns (Fig. 2.3H: II): direct loops (blue) comprised only a small fraction of total loops and were significantly smaller than indirect loops (green), consistent with genome-wide observations (Fig. 2.3A). In contrast, GAF-dependent loops exhibited distinct characteristics (Fig. 2.3H: IV): direct GAF loops showed greater size diversity and more engagement with diverse loci, mirroring the broader size distribution observed at HAS (Fig. 2.3D, right). Notably, TF occupancy did not always predict direct looping: 66.3% of CLAMP ChIP-seq peaks (N=8,305/12,518) and 55.4% of GAF peaks (N=6,465/11,680) were not associated with Hi-ChIP loops (Fig. 2.3H: II, IV). These findings indicate that CLAMP and GAF occupancy and direct looping are not always coupled, potentially due to redundant architectural proteins that maintain loops following TF depletion and/or additional functions of CLAMP and GAF beyond chromatin looping.

Notably, the disruption of CLAMP- and GAF-dependent looping resulted in differential TAD formation under both *clamp* and *gaf* RNAi knockdown conditions (Fig. 2.3H: I, III; Fig. S2.3). Given that HAS (Fig. 2.3H, orange highlighted loci) are known to populate TAD boundaries (Ramírez et al. 2015), these findings reinforce the role of these factors in establishing and maintaining TAD integrity. To quantify the relationship between loop loss and TAD reorganization, we compared the differential TADs (Fig. 2.3H: I, III) with our direct loop sets (Fig. 2.3H: II, IV), determining what proportion of TAD structural changes could be directly attributed to the loss of CLAMP or GAF loops. Direct CLAMP-dependent loops were associated with 41.7% (N=508/1218) of differential CLAMP TADs, while direct GAF-dependent loops were associated with 32.4% (N=1105/3415) of differential GAF TADs. This substantial overlap between direct loop loss and TAD reorganization provides quantitative evidence that CLAMP and GAF play direct, mechanistic roles in maintaining TAD architecture, rather than acting solely through indirect or compensatory mechanisms.

Lastly, direct CLAMP- and GAF-dependent looping affect a variety of genomic elements as seen by the diverse chromatin states and genes enriched within a specific region (Fig. 2.3H: V, VI). Therefore, we hypothesized that it may be that these distinct, direct looping contacts regulate interactions between different chromatin states on the X chromosome compared with autosomes.

2.3.4 GAF regulates chromatin looping at silent euchromatin specifically on the X chromosome and not on autosomes while CLAMP regulates chromatin looping between active chromatin regions throughout the genome

To investigate the distinct functions of CLAMP and GAF in directly regulating chromatin looping on the X chromosome and autosomes, we analyzed sequence elements and chromatin states that are present at directly regulated loop anchors (Fig. 2.4). Interestingly, we observed that while GAF-mediated loops on autosomes predominantly connect active chromatin states (States 1 and 3), GAF predominantly regulates looping at inactive silent euchromatin on the X chromosome (State 9) (Fig. 2.4A, B). Our data suggest that the function of the GAF-CLAMP competition (Kaye et al. 2018) may be to alter the location of chromatin loops directly regulated by GAF specifically on the X chromosome which promotes CLAMP and not GAF occupancy at active chromatin regions that require dosage compensation (Fig. 2.3E-H).

While GAF regulates the formation of chromatin loops between active chromatin states rather than silent regions on autosomes, on the X chromosome, we observed GAF often tethers active transcriptional chromatin states (States 1, 3, and 5) to transcriptionally silent states (State 9) (Fig. 2.4A, B). Notably, the loop anchors at silent states on the X chromosome, where chromatin looping is directly regulated by GAF, are located within transcriptionally silent intergenic euchromatin and insulator-rich regions (State 9) rather than Polycomb-repressed (State 6) or constitutive heterochromatin (States 7 and 8) (modENCODE Consortium et al. 2010). Therefore, GAF regulates looping that brings DC genes near insulators but not heterochromatin (Fig. 2.4A, B). We also found that direct GAF-mediated loops have a larger median loop size on the X chromosome (8kb–35kb), where GAF is outcompeted by CLAMP, than on autosomes (9-13kb) (Fig. 2.4A, B) (Kaye et al. 2018). Overall, chromatin loops that are regulated by GAF on the X chromosome differ significantly from those on autosomes because they are enriched for direct links between insulators

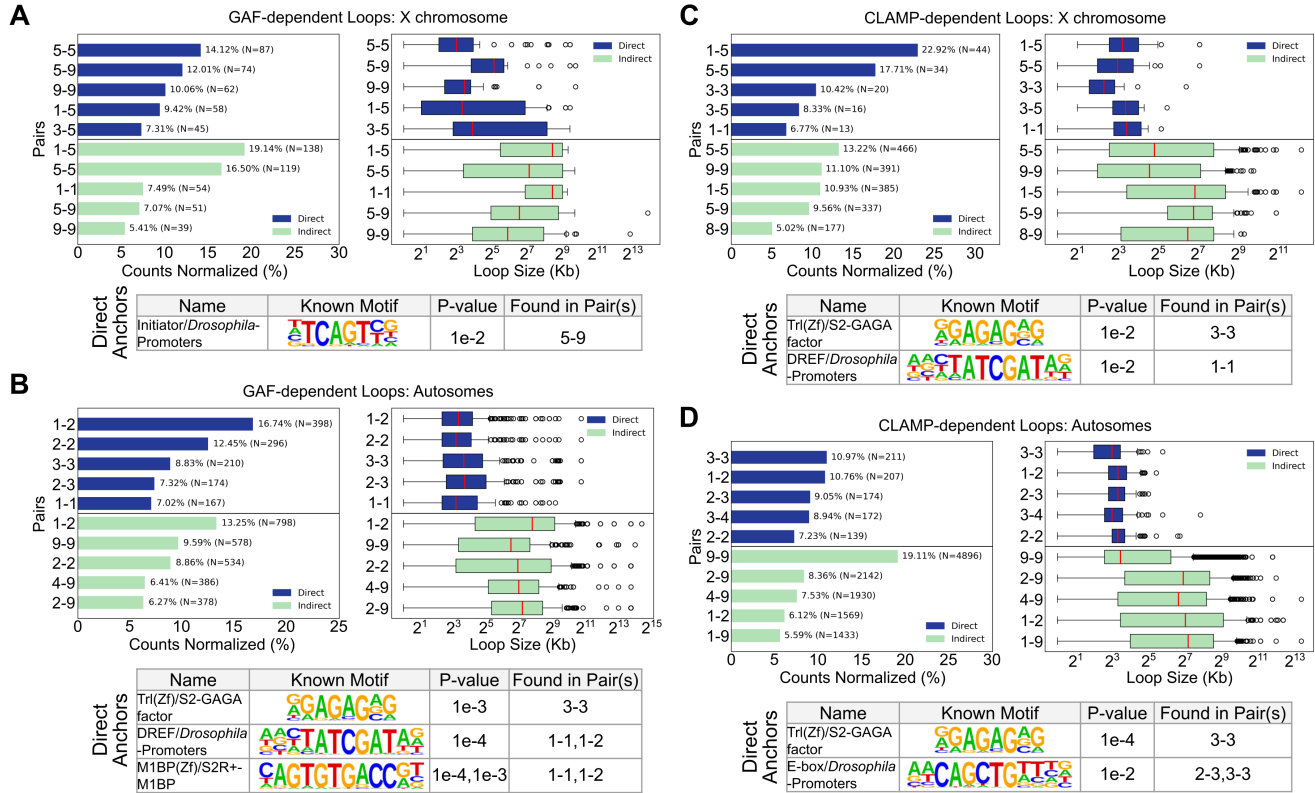


Figure 2.4: Chromatin State Pair enrichment of CLAMP- and GAF-dependent loops shows GAF regulates chromatin looping at silent euchromatin on the X chromosome and not on autosomes. CLAMP on the (A) X chromosome and (B) autosomes: State Pair Enrichment: Normalized counts (%) of direct (blue) and indirect (green) down-regulated loops, categorized by the chromatin state pairs at their anchors. Loop Size (Top): Comparison of genomic span (kb) for direct and indirect loops across enriched state pairs. On the X chromosome, direct loops frequently bridge State 1 and 3 (Active Promoters and Enhancers) with State 5 (DC genes). Size Distribution: Direct interactions consistently exhibit a smaller, more constrained size distribution compared to the larger span of indirect contacts. Motif Enrichment (Bottom): Known motif discovery analysis for loop anchor pairs. Significant motifs include the GAGA-factor (*Trl*) and DREF promoter elements specifically enriched at direct CLAMP-dependent anchors. (C-D) GAF-mediated chromatin looping. GAF on the (C) X chromosome and (D) autosomes: State Pair Enrichment: Distribution of direct and indirect GAF-dependent loops across chromatin state pairs. GAF preferentially organizes loops between State 9 (transcriptionally silent euchromatin) and State 5 (DC genes) on the X chromosome, while facilitating active chromatin contacts (States 1-3) on autosomes. Loop Size (Top): Comparison of genomic span (kb) for GAF-mediated interactions. Direct GAF loops (blue) maintain a larger genomic span on the X chromosome than on autosomes and indirect GAF loops (green) maintain a broader genomic span across all categories. Motif Enrichment (Bottom): Known motif discovery analysis for loop anchor pairs at direct GAF-mediated anchors shows enrichment of Initiator (*Inr*) motif on the X chromosome while GAGA-factor (*Trl*) and DREF motifs were enriched on autosomes.

located within silent euchromatin and other genomic loci (Fig. 2.4A, B).

In contrast to GAF, CLAMP directly mediates chromatin looping between distal enhancers (State 3) and promoters (State 1) and DC genes on the X chromosome (State 5) (Fig. 2.4C). On autosomes, CLAMP direct loops also bridge active chromatin states (Fig. 2.4D). The median loop size of direct CLAMP loops is 8-11kb on the X chromosome and 8-10kb on autosomes, suggesting that CLAMP regulates loops of similar sizes on the X chromosome and autosomes (Fig. 2.4C, D). Therefore, CLAMP directly facilitates precise, short-range chromatin interactions between transcriptionally active chromatin and dosage-compensated genes on the X chromosome and between active genes on autosomes.

Distinguishing the direct from indirect effects of GAF and CLAMP is critical because direct and indirect chromatin loops occur at very different chromatin states. While the chromatin loops that are directly facilitated by CLAMP and GAF are distinct from each other in scale and chromatin state, the majority of chromatin loops that change after the loss of either factor are indirectly regulated (Fig. 2.3B, C). Furthermore, there are differences between the loops that are indirectly regulated by CLAMP and GAF: we observed that while CLAMP directly mediates chromatin loops only between loci within active states (States 1, 3, and 5), loops indirectly regulated by CLAMP include inactive states (States 8 and 9) (Fig. 2.4C). Conversely, loops indirectly regulated by GAF include an active state (State 1) on the X chromosome which was not observed in the directly regulated chromatin loops (Fig. 2.4A). Therefore, it is key to distinguish direct from indirect effects to understand the differential functions of CLAMP and GAF. Future analysis will define how the loss of CLAMP or GAF indirectly regulates chromatin looping through modulating the function of other looping factors.

To identify potential cofactors that may function with CLAMP or GAF, we performed HOMER motif analysis on direct CLAMP and GAF loop anchors genome-wide. As expected, we find the Trl(Zf)/GAGA-factor motif ($P = 1e-4$; Fig. 2.4D) at CLAMP-mediated 3-3 loop anchors on X ($P = 1e-2$; Fig. 2.4C) and autosomes ($P = 1e-4$; Fig. 2.4D), likely due to the high sequence similarity between CLAMP and GAF (Fig. 2.3F). Additionally, at CLAMP-mediated loops on the X chromosome, we found the DNA replication-related element factor (DREF) binding motif (Fig. 2.4C), a factor which binds and activates housekeeping genes (Zabidi et al. 2015). The enrichment of the DREF motif at promoter-promoter (1-1) contacts ($P = 1e-2$; Fig. 2.4C) suggest that loop anchors that are directly regulated by CLAMP are enriched for housekeeping genes on the X chromo-

some, which are the first genes to be compensated during development (Prayitno et al. 2019). In contrast, on autosomes, CLAMP regulates chromatin loop anchors enriched for E-box enhancer elements at enhancer–enhancer (3-3) and exon–enhancer (2-3) contacts ($P = 1e-2$; Fig. 2.4D), where developmental genes would require canonical E-box on-off switches (Gordân et al. 2013).

We next examined motifs associated with chromatin loops directly regulated by GAF. GAF-mediated loops on the X chromosome that connect dosage compensation genes (state 5) with silent euchromatin insulator regions (State 9) are enriched for the Initiator (*Inr*) motif (Fig. 2.4A). *Inr* is a promoter element that often cooperates other core promoter elements as it cannot initiate transcription alone (Vo Ngoc et al. 2019). It is possible that at proximal promoter regions of DC genes, GAF may require additional factors, such as co-promoter element *Inr*, to facilitate looping on the X chromosome.

Unlike the X-chromosome loop anchors, autosomal GAF-mediated loop anchors are enriched for motifs exclusively within active chromatin states (States 1-3) (Fig. 2.4B). The presence of DREF and Motif-1-binding protein (M1BP), an RNA Pol II pausing factor associated with housekeeping genes (Poliacikova et al. 2023), at active promoter–exon (1-3) contacts suggests a distinct autosomal role for GAF (Fig. 2.4B). On autosomes, GAF may facilitate looping to regulate housekeeping genes consistent with prior work (Li and Gilmour 2013). Overall, CLAMP direct loop anchors are at housekeeping genes on the X chromosome that need to be dosage-compensated, and GAF direct loop anchors are at housekeeping genes on autosomes suggesting that the CLAMP-GAF competition may allow CLAMP to be enriched at X-linked housekeeping genes instead of GAF to allow these genes to recruit the DCC.

We next compared the size of different classes of loops connecting different chromatin states on the X chromosome and autosomes (Fig. 2.4). We observed that on the X chromosome, GAF-mediated state 5 (dosage compensation)-state 9 (insulators) loops are significantly larger than state 5-5 ($P = 3.83e-8$) or state 9-9 ($P = 1.94e-8$) loops (Fig. 2.4A). Additionally, state 9-9 loops were significantly smaller than state 3-5 loops ($P = 1.84e-2$; Fig. 2.4A). Therefore, it is possible that GAF mediates larger state 5-9 chromatin loops that connect insulators to dosage compensated genes to provide a three-dimensional scaffold for the dosage compensation molecular hub. This larger GAF-mediated loop scaffold may explain the function of GAF in dosage compensation that is weaker than that of CLAMP but still important for DCC recruitment (Kaye et al. 2018).

Collectively, our results reveal that CLAMP and GAF deploy context-dependent strategies to organize chromatin architecture: CLAMP directly mediates short-range loops between active chromatin states on both X and autosomes. In contrast, the function of GAF diverges based on chromosomal context, facilitating compact loops between active loci on autosomes, but forming larger loops between dosage-compensated genes and silent euchromatin on the X chromosome. These distinct looping patterns reflect the competitive dynamics between CLAMP and GAF on the X chromosome, where the exclusion of GAF from active sites by CLAMP likely redirects it to silent regions where it regulates larger loops that may provide structural scaffolding for DCC spreading. Our findings demonstrate how competitive binding enables CLAMP and GAF to context-specifically regulate chromatin loops that are distinct in both size and chromatin state.

2.3.5 Chromatin loops directly mediated by CLAMP and GAF associate with *roX2* chromatin loops on the X chromosome

The DCC incorporates one of two long non-coding RNAs (lncRNAs), *roX1* or *roX2*, and therefore these lncRNAs are spatial markers of DCC localization (Meller and Rattner 2002). Tian et al. (2024) performed the RNA-associated chromatin DNA-DNA interaction (RDD) method targeting *roX2* across the genome in *Drosophila* S2 cells to capture chromatin loops associated with this lncRNA and observed the expected X-chromosome specificity. Furthermore, we recently demonstrated that CLAMP can directly interact with *roX2* in vivo and in vitro (Ray, M*, Zaborowsky, J*, Mahableshwarkar, P, Aguilera, J, Huang, A, Vaidyanathan, S, Fawzi, N and Larschan, E. 2024). We therefore leveraged these *roX2*-associated loops as a proxy for three-dimensional DCC localization and spreading across the X chromosome, enabling us to determine the proportion of direct CLAMP- and GAF-dependent loops that spatially associate with the DCC beyond the HAS. Prior to integrating our differential loop datasets with the *roX2*-RDD chromatin loops, we equalized bin sizes across all datasets to enable direct comparison (see Methods), and we report the complete loop counts after normalization (Table S2.2).

We first asked what proportion of *roX2*-RDD loops on the X chromosome overlap with direct CLAMP- and GAF-dependent loops, and whether this proportion is enriched compared to indirect loops. For CLAMP-dependent loops, direct loops overlapped with *roX2*-RDD loops significantly more than their corresponding indirect loops at HAS (5.2% vs 2.0%; $P < 0.05$, Fisher's Exact Test)

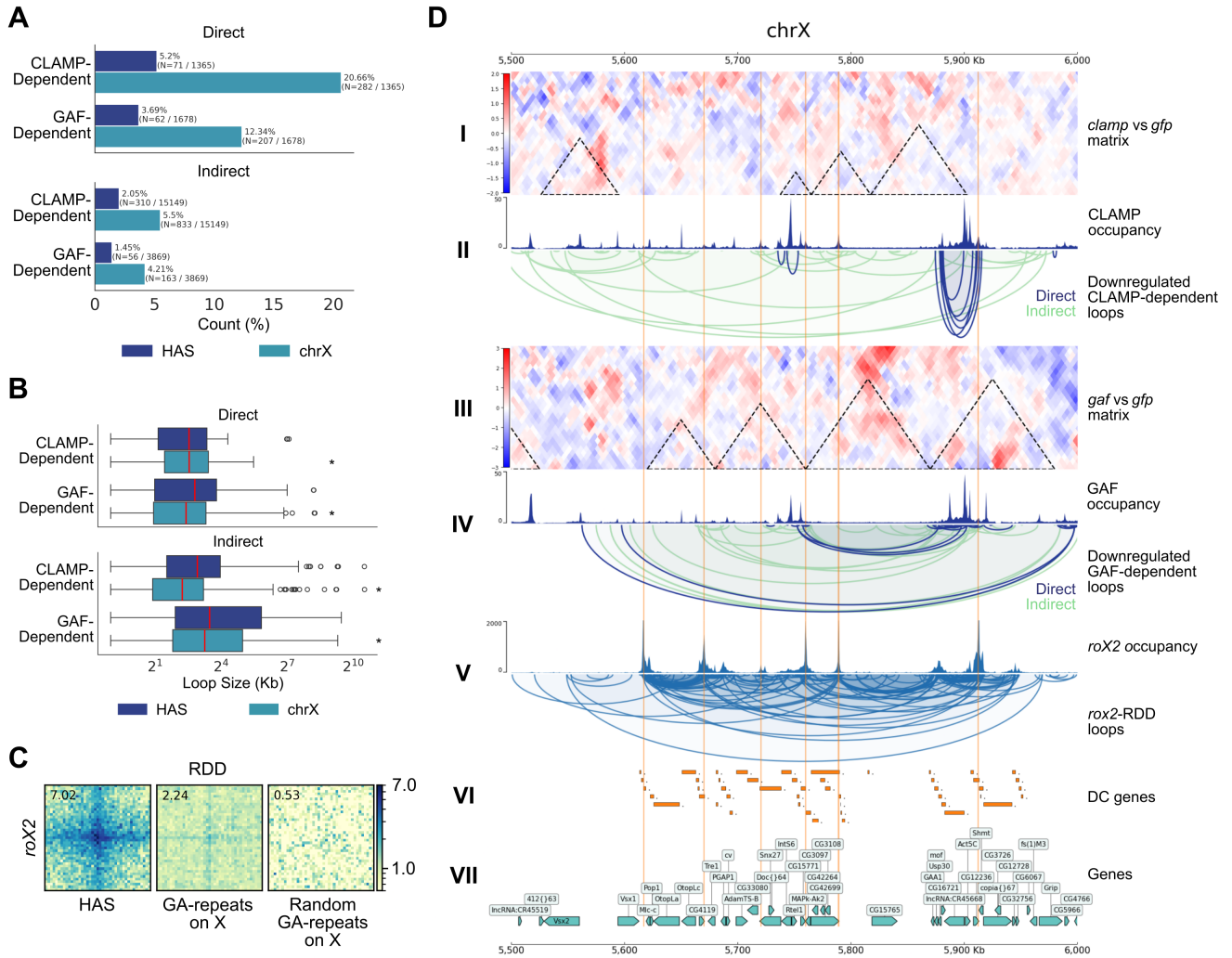


Figure 2.5: Direct CLAMP- and GAF-dependent loops are associated with a subset of roX2 chromatin loops on the X chromosome. (A) Distribution of direct and indirect loops associated with roX2 loops at HAS and across the X chromosome. Bar plots showing the normalized counts (%) of direct and indirect CLAMP- and GAF-dependent loops that overlap with roX2 loops at High Affinity Sites (HAS; dark blue) compared to the X chromosome (chrX; cyan). (B) Loop size of roX2-associated loops. Box plots comparing the loop size (kb) for direct and indirect CLAMP- and GAF-dependent interactions at HAS versus chrX. (C) Aggregate Peak Analysis (APA) of roX2 loops over HAS. APA plots visualizing roX2 RNA-associated chromatin DNA-DNA (RDD) interactions at HAS, GA-repeats on the X, and random GA-repeats on the X. Strong enrichment at HAS (7.02) shows that roX2 binds preferentially at HAS. Pixel resolution is 1.0kb with ± 20 kb flanking central pixel. (D) Co-localization of CLAMP- and GAF-dependent loops with roX2-RDD loops. Representative CoolBox plots of a 500kb genomic window shows (I & III) differential contact matrices (log2FC) for *clamp* RNAi and *gaf* RNAi versus *gfp* RNAi control, respectively. (II & IV) Hi-ChIP CLAMP and GAF protein occupancy tracks aligned with downregulated direct and indirect chromatin loops. (V) roX2 occupancy aligned with roX2-RDD loops. (VI) Genomic tracks for Dosage Compensation (DC) genes and (VII) dm6 gene annotations on the X chromosome.

and on the X chromosome (20.7% vs 5.5%; $P < 0.05$, Fisher's Exact Test) (Fig. 2.5A). The same pattern was observed for GAF-dependent loops, with direct loops similarly enriched compared with indirect loops at HAS (3.7% vs 1.5%; $P < 0.05$, Fisher's Exact Test) and on the X chromosome (12.3% vs 4.2%; $P < 0.05$, Fisher's Exact Test) (Fig. 2.5A). These results suggest that under normal conditions where dosage compensation is active, *roX2*-RDD loops at HAS and on the X chromosome spatially associate with direct CLAMP- and GAF-dependent regulation, linking these factors spatially to the DCC. However, the majority of loops in each dataset do not overlap which is consistent with *roX2* spreading to active gene bodies where CLAMP and GAF largely do not localize, and the well-characterized functions of both TFs beyond dosage compensation.

Next, given our earlier observation of TF-specific loop size distributions (Figs. 2.3, 2.4), we asked whether the *roX2*-associated direct loops exhibit distinct size characteristics compared to indirect loops. On the X chromosome, the loop size distributions of direct and indirect sets were significantly different for both CLAMP- and GAF-dependent loops ($P < 0.05$, Mann-Whitney Test), though this difference was not significant at HAS (Fig. 2.5B). Specifically, direct CLAMP- and GAF-dependent loops both exhibited a narrower range of loop sizes than indirect loops (Fig. 2.5B), consistent with the local looping patterns observed previously (Fig. 2.3). The preferential overlap of direct CLAMP- and GAF-dependent loops with *roX2*-RDD loops, combined with their shorter-range size profiles and reduced overlap at HAS, suggests that these factors may contribute to the local spreading of dosage compensation on the X chromosome.

Having established the unique CLAMP- and GAF-specific looping characteristics at HAS and GA-repeats (Fig. 2.3), we next asked how *roX2*-RDD loops aggregate over these same classes of sites on the X chromosome (Fig. 2.5C). Using APA plots, we observed strong aggregated *roX2*-RDD looping at HAS, recapitulating the central role of this lncRNA and the DCC in X-chromosome dosage compensation (Fig. 2.5C). We then examined *roX2*-RDD looping over GA-repeats that directly associate with CLAMP and GAF on the X chromosome. Notably, aggregated *roX2*-RDD looping and direct CLAMP-dependent looping at both HAS and across the X chromosome exhibited a similar striped signal pattern, indicating that *roX2*-RDD loops, like direct CLAMP-dependent loops, favor broad, directional chromatin interactions emanating from GA-rich loci rather than forming discrete point-to-point contacts (Fig. 2.3F, 2.5C). However, aggregated *roX2*-RDD looping at HAS (log enrichment = 7.02) was substantially greater than that of either direct CLAMP- or GAF-

dependent loops (log enrichment = 1.80 and 1.68 respectively) (Fig. 2.5C). These results suggest that HAS serve as key regulatory hubs for dosage compensation, and that *roX2*-associated looping more closely resembles direct CLAMP-dependent looping than GAF-dependent looping, consistent with our previous work demonstrating how the DCC (which includes *roX2*) and CLAMP may cooperate at HAS to mediate the local spreading of dosage compensation on the X chromosome (Ray, M*, Zaborowsky, J*, Mahableshwarkar, P, Aguilera, J, Huang, A, Vaidyanathan, S, Fawzi, N and Larschan, E. 2024; Soruco et al. 2013).

Lastly, to visualize *roX2*-RDD loops alongside direct CLAMP- and GAF-dependent loops, we examined a representative region on the X chromosome (chrX:5,500,000-6,000,000) (Fig. 2.5D). Within this region, differential CLAMP- and GAF-dependent loops exhibited looping characteristics consistent with genome-wide patterns (Fig. 2.3D, Fig. 2.5D: II, IV). The *roX2*-RDD loops encompassed a broad size distribution at this locus (Fig. 2.5D: V), a pattern also observed genome-wide (Fig. SS2.4), spanning both intra- and inter-TAD interactions (Fig. 2.5D: I, III, V). Notably, *roX2* and CLAMP occupancy were enriched at HAS (orange highlighted loci), while GAF occupancy was not (Fig. 2.5D: II, IV, V), further supporting the closer functional relationship between *roX2* and CLAMP at these sites compared with GAF. In addition, *roX2*-RDD loops were present at every control TAD on the X chromosome, suggesting that *roX2*-associated looping is a ubiquitous feature of X-chromosome architecture essential for dosage compensation. These *roX2*-RDD loops also encompassed all genes undergoing dosage compensation (Fig. SS2.5, Fig. 2.5D: VI, VII), consistent with a direct spatial relationship between *roX2*-associated chromatin architecture and the transcriptional output of the DCC. Together, we demonstrate that *roX2*-RDD loops span the X chromosome and spatially associate with direct CLAMP- and GAF-dependent loops, illustrating how these factor-specific loops shape the three-dimensional context for dosage compensation.

2.4 Discussion

A fundamental question about gene regulation across species is: How are the correct genes targeted for coordinate regulation within the highly compact three-dimensional nucleus? We used the upregulation of X-linked genes (XCU) as a model to address this question because we have identified key DNA sequences (Alekseyenko et al. 2008, 2012) and transcription factors (TFs)

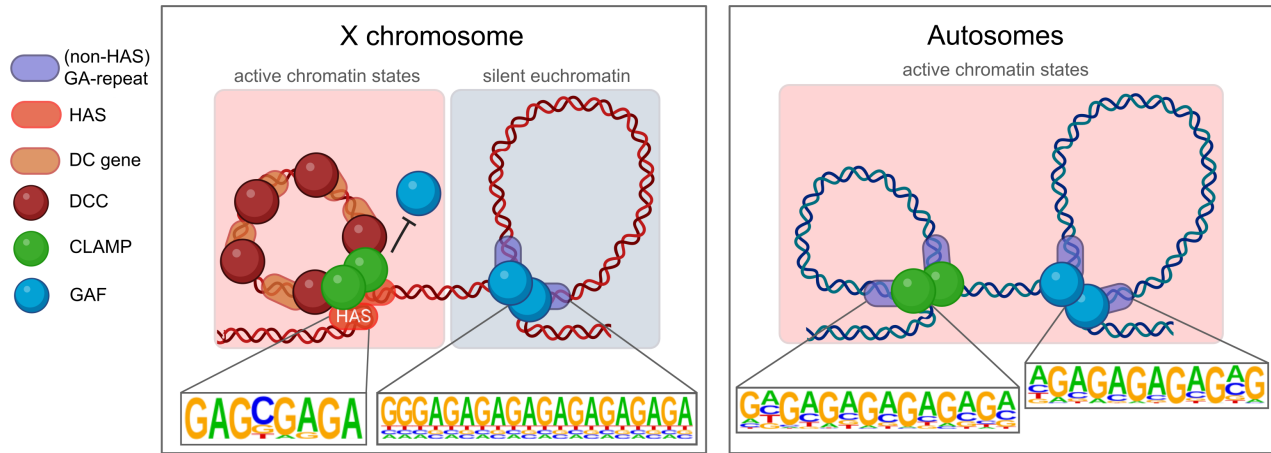


Figure 2.6: **Model for X-specific CLAMP and GAF function in chromatin looping.** (1) On the X chromosome, a single base change within GA-rich TF binding sites alters the relative affinity of CLAMP and GAF which then favors the formation of smaller CLAMP driven loops at active dosage-compensated housekeeping genes and larger GAF-driven loops at silent insulator regions. (2) In contrast, on autosomes both CLAMP and GAF mediate loops between active genes, although GAF-regulated loops have a broader range of sizes than CLAMP-regulated loops.

(Sorucu et al. 2013; Larschan et al. 2012) that are involved in targeting the DCC specifically to the X-chromosome, including competition between two essential pioneer TFs (Kaye et al. 2018). However, the mechanisms by which sequence elements and TFs function directly to establish an X-specific three-dimensional chromatin environment remained unknown. Here, we provide mechanistic evidence to support the following model for X-chromosome specificity (Fig. 2.6): 1) On the X-chromosome, a single base change within GA-rich TF binding sites alters the relative affinity of CLAMP and GAF which then favors the formation of smaller CLAMP driven loops at active dosage-compensated housekeeping genes and larger GAF-driven loops at silent insulator regions. 2) In contrast, on autosomes both CLAMP and GAF mediate loops between active genes, although GAF-regulated loops have a broader range of sizes than CLAMP-regulated loops. Overall, we provide insight into how the X chromosome is recognized for dosage compensation.

There are many contexts in which the DNA sequence content of TF binding sites varies, including within SNPs related to disease (Yan et al. 2021; Khetan et al. 2025). Our data suggest that even single base alterations of TF binding sites can alter three-dimensional nuclear organization by favoring the binding of one TF instead of another which can drive different transcriptional programs. For example, the accumulation of binding sites on the X chromosome that favor CLAMP and not GAF near housekeeping genes that need to be dosage compensated promotes an X-specific three-dimensional organizational hub that allows CLAMP to recruit the DCC to genes that

need to be dosage compensated.

GAF also contributes to DCC recruitment but to a much lesser extent than CLAMP (Kaye et al. 2018). Based on our three-dimensional analysis of the direct function of GAF, it is possible that GAF contributes to DCC recruitment by generating an X-specific three-dimensional scaffold consisting of larger chromatin loops between X-linked insulators. In contrast, on autosomes GAF regulates interactions between active chromatin regions, such as enhancers and promoters. This context-specific architectural loop function is consistent with prior reports that suggest that GAF is an important component of multiple insulator complexes (Lomaev et al. 2017; Wolle et al. 2015) and has a function in dosage compensation (Kaye et al. 2018, 2017; Greenberg et al. 2004). Furthermore, GAF also mediates several very long-range looping interactions in the brain that are important for regulating functionally similar genes that are encoded at different locations in the genome (Li et al. 2023). Therefore, our data provide mechanistic insight to explain prior observations that GAF functions as an insulator protein and has a role in dosage compensation.

In contrast to the function of GAF at insulator loci on the X chromosome where it facilitates long-range interactions, we demonstrate that CLAMP facilitates small scale loops between active regions on both the X chromosome and autosomes. It is possible that these smaller hubs are critical for the spreading of the DCC from the HAS to the bodies of active genes where it increases transcript levels. Interestingly, the first genes to be compensated that are in close two-dimensional proximity to the HAS are the housekeeping genes which encode components of essential protein complexes that require precise stoichiometry (Prayitno et al. 2019). Housekeeping genes are often enriched for the DREF motif (Hochheimer et al. 2002) which we found as highly enriched at loops which were directly mediated by CLAMP but not GAF on the X-chromosome and at autosomal loops regulated by GAF and not CLAMP. This inverse relationship between looping at housekeeping genes on the X chromosome and autosomes suggests that competition between CLAMP and GAF is important to assure that CLAMP but not GAF is enriched at X-linked housekeeping genes that require dosage compensation.

There are many possible biomolecular mechanisms by which CLAMP and GAF could regulate mutually exclusive sets of looping interactions. One possibility is that their intrinsically disordered domains (IDRs) that do not bind DNA have an independent set of interactors and biophysical properties. For example, the CLAMP IDR interacts with both the *roX* RNAs (*in vivo* and *in vitro*)

(Ray, M*, Zaborowsky, J*, Mahableshwarkar, P, Aguilera, J, Huang, A, Vaidyanathan, S, Fawzi, N and Larschan, E. 2024) and MSL2 (Eggers et al. 2023; Valsecchi et al. 2020; Tikhonova et al. 2019) which are not known GAF interactors. Furthermore, our work is consistent with a very recent study that indicates that heterotypic interactions between a wide variety of sequence motifs including the GA-rich motif are important for boundary formation in *Drosophila* (Varisco et al. 2025).

The function of loop extrusion in defining three-dimensional organization in *Drosophila* had been debated until a very recent study quantitatively defined the function of loop extrusion compared to that of tethering elements which can be regulated by GAF (Choppakatla et al. 2026). Tethering elements and loop extrusion are both important for chromatin loop formation, but they have different roles. Loop extrusion functions more strongly during the process by which enhancers and promoters are searching for the correct contacts whereas tethering elements function to stabilize these contacts and regulate transcriptional output. Our work suggests that the specific TFs that bind to tethering elements control the size and chromatin context of the loops that are stabilized. Overall, we provide mechanistic insight into how specific genomic loci are targeted for coordinated gene regulation within a highly compacted three-dimensional chromatin environment.

2.5 Methods

2.5.1 Cell Culture

S2 cells were started from frozen stocks obtained from the *Drosophila* Genomics Resource Center and maintained at 25°C in Schneider's media supplemented with 10% Fetal Bovine Serum and 1.4X Antibiotic-Antimycotic (ThermoFisher Scientific, USA). Cells were passaged every 3 days to maintain an appropriate cell density.

2.5.2 Western Blot

Protein was extracted from cell pellets using Cell Lysis Buffer (Cell Signaling Technology, 9803) supplemented with 1 mM protease inhibitor cocktail (Abcam, ab65621). Protein concentrations were determined using the DC Protein Assay (Bio-Rad, 5000116). Equal amounts of protein were mixed with NuPAGE™ Sample Reducing Agent (Thermo Fisher Scientific, NP009) and NuPAGE™ LDS Sample Buffer (Thermo Fisher Scientific, NP0007), boiled at 70°C for 10 min, and loaded onto

a 4%–12% gradient SurePAGE™ Bis-Tris gel (GenScript, M00652). Proteins were transferred to a methanol-activated PVDF membrane using a semi-dry transfer method with the Trans-Blot Turbo RTA Transfer Kit PVDF (Bio-Rad, 1704273), Trans-Blot Turbo 5× Transfer Buffer (Bio-Rad, 10026938), and the Bio-Rad Trans-Blot Turbo Transfer System (1.3A, 25 V) for 10 min. Membranes were blocked in 5% milk in phosphate-buffered saline containing 0.05% Tween-20 (PBST) for 20 min at room temperature, then incubated with primary antibodies diluted in 5% milk in PBST overnight at 4°C. Membranes were washed with PBST and incubated with HRP-conjugated secondary antibodies diluted in 5% milk in PBST for 1 h at room temperature, followed by additional washes in PBST. Protein bands were detected using SuperSignal™ West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific, 34095) and imaged with the Bio-Rad ChemiDoc Imaging System. GAPDH was used as a loading control and was detected using an HRP-conjugated GAPDH monoclonal antibody. All blots are shown in the Supplementary Material. Antibodies and dilutions were as follows: GAPDH (Proteintech, HRP-60004, 1:1000), CLAMP (1:1000), GAF (1:1000), Anti-Rabbit (Cell Signaling, 7074S, 1:1000).

2.5.3 Micro-C

Micro-C was performed on S2 cells using two biological replicates for each of the following conditions: *clamp* RNAi, *gaf* RNAi, and *gfp* RNAi (control). For each replicate, 3×10^6 cells were incubated with 15 µg of the corresponding dsRNA, resulting in six samples total. S2 cells were treated in 6-well plates with 15 µg of dsRNA (*clamp*, *gaf*, or *gfp* control) in 500 µl of water. S2 cells were first incubated with the dsRNA in Schneider's media without FBS or Antibiotic-Antimycotic for 45 minutes. After 45 minutes, 3 mL of medium supplemented with 10% FBS and Antibiotic-Antimycotic was added to each well. The plates were then incubated at 25°C for six days before harvesting. After incubation, cells in each well were scraped, aliquoted into 1.5 ml tubes, then spun at 500g in a swinging bucket rotor for 5 minutes to pellet the cells. Supernatant was removed and cell pellets were washed with 1x PBS, counted, then 3×10^6 cells (protocol-recommended cell count of 1×10^6 was optimized to account for reduced size of *Drosophila* cells) were re-pelleted in 1.5ml tubes and placed in -80°C for 1 hour as per the manufacturer's protocol (Dovetail Micro-C Kit user guide version 2.0). To prepare samples, cells were thawed at RT and resuspended in 200 µL of 1x PBS (ThermoFisher Scientific, 10010023) and 2 µL 0.3 M DSG. 37% formaldehyde was then

added after a 10-minute incubation on a tube rotator to complete crosslinking, as per the manufacturer's protocol (Dovetail Micro-C Kit user guide version 2.0). Sample digestion, quantification (Tapestation HS D5000), proximity ligation, library preparation, ligation capture, and amplification were done using the Dovetail Micro-C Kit (Catalog #21006), the Dovetail Library Module for Illumina (Catalog #25004), and the primer set for NGS library preparation (Dovetail Dual Index Primer Set #1 for Illumina, Catalog #25010). Libraries were sequenced with Illumina 150x2 bp pair-end sequencing at 80X coverage. Data was submitted to Gene Expression Omnibus (GSE322508).

2.5.4 *Hi-ChIP*

S2 Cells were allowed to grow to confluence then harvested. Two replicates of 10×10^6 S2 cells were used per HiChIP reaction. Cells were crosslinked using 0.3M DSG and 37% formaldehyde as per the manufacturer's protocol (Dovetail HiChIP MNase Kit user guide). Digestion, lysate preparation, chromatin immunoprecipitation, proximity ligation, and library preparation was conducted using the Dovetail HiChIP MNase Kit (Catalog # 21007) followed by the Dovetail Library Prep Kit and the primer set for NGS library preparation (included in kit). For CLAMP and GAF immunoprecipitation, Rabbit anti-CLAMP (5µg/sample, SDIX) and Rabbit anti-GAF (5µg/sample) were used, respectively. Libraries were subjected to Illumina 150x2 bp pair-end sequencing. Data was submitted to Gene Expression Omnibus (GSE322509).

2.5.5 *Micro-C Preprocessing*

Micro-C raw data was analyzed using the HiC-Pro toolkit (v3.1.0) (Servant et al. 2015) with default parameters. Normalized contact matrices deriving from HiC-Pro were converted to the cooler format using utility scripts (<https://github.com/nservant/HiC-Pro/tree/master/bin/utils>) in the HiC-Pro toolkit (v3.1.0) and replicates were subsequently merged.

2.5.6 *Hi-ChIP Preprocessing*

HiChIP raw data was analyzed using the Dovetail genomics Hi-ChIP documentation (<https://hichip.readthedocs.io/en/latest/index.html>). Sequences were mapped to the dm6 genome using the Burrows-Wheeler Aligner (v0.7.17) (Li and Durbin 2009; Li 2013) passing the following arguments: mem,

-5, -S, -P, -T0; default values were used for all other settings. Using the parse module of pairtools (v0.3.0) (Open2C et al. 2024), valid ligation events were recorded; the following parameters were changed from default values: `--min-mapq 40`, `--walks-policy 5unique`, `--max-inter-align-gap 30`. PCR duplicates were removed using the dedup module of pairtools (v0.3.0) (Open2C et al. 2024). Subsequently, the split module of pairtools was used to generate the final bam and pair files. The samtools (v1.17) (Danecek et al. 2021) sort module was used to sort the bam files. To determine the quality of the proximity ligation events, scripts generated by the Dovetail Genomics team were utilized, which are deposited in the following GitHub repository (<https://github.com/dovetail-genomics/HiChIP>).

The cooler suite (v0.9.3) (Abdennur and Mirny 2020) was used to generate contact maps from pair files and replicates were subsequently merged. To call loops, the flexible FitHiChIP tool (v10.0) (Bhattacharyya et al. 2019) was used as resolutions are easily adjustable, and it can detect different categories of 3D chromatin interactions. Furthermore, since FitHiChIP required 1D peaks from the same tissue and conditions, MACS2 (v2.2.9.1) (Zhang et al. 2008) was used to call peaks on the Hi-ChIP data. Consensus loops were created for each condition and were used throughout the study.

2.5.7 Computational Analysis

Topologically-associated domain (TAD) calling and differential analysis (Micro-C)

TADs were identified for all Micro-C conditions (*gfp*, *clamp*, and *gaf* RNAi) using HiCEXplorer (v3.7.5) (Wolff et al. 2020, 2018; Ramírez et al. 2018). TADs were called at six resolutions (2.5kb, 5kb, 10kb, 20kb, 40kb, and 80kb) using the hicFindTADs tool with the following parameters: `--minDepth RESOLUTION × 3`, `--maxDepth RESOLUTION × 10`, `--step RESOLUTION`, and `--correctForMultipleTesting fdr`. Domains across all resolutions were then merged using hicMergeDomains with default parameters to generate multi-resolution consensus TAD sets for each RNAi condition.

To assess how TAD structure was altered upon knockdown of CLAMP or GAF, we used the *gfp* RNAi condition as the reference and performed differential TAD analysis using the following HiCEXplorer (v3.7.5) (Wolff et al. 2020, 2018; Ramírez et al. 2018) workflow: (1) hicNormalize

(`--normalize smallest`; default parameters) to normalize contact matrices across conditions; (2) `hicCorrectMatrix` (default parameters) to remove systematic biases via iterative correction (ICE); and (3) `hicDifferentialTAD` (`--targetMatrix TREATMENT-MATRIX`, `--controlMatrix CONTROL-MATRIX`, `--tadDomains CONTROL-TADS`, `--pValue 0.05`, `--mode all`; default parameters) to identify differential TADs relative to the *gfp* RNAi reference. Differential TAD analysis was performed at each of the six resolutions described above, and results were subsequently merged.

Multi-resolution differential chromatin loop analysis (Micro-C)

Merged cooler contact maps that were produced for each condition were used to identify differential chromatin interactions using the HiCcompare library (v1.24.0) (Stansfield et al. 2018) in Bioconductor (v3.18) (Huber et al. 2015). Differential chromatin interactions were determined at three different resolutions: 1kb, 5kb, and 10kb. Consensus differential chromatin interactions were constructed by merging all three resolutions using a custom Python script that followed the `hicMergeLoops` method in HiCExplorer (v3.7.5) (Wolff et al. 2020, 2018; Ramírez et al. 2018). Consensus significant differential chromatin interactions ($P < 0.01$) were selected and used throughout the study.

TAD and differential loop integration (Micro-C)

Differential TADs were integrated with the consensus merged sets of differential chromatin interactions using the pybedtools suite (v0.11.0) (Dale et al. 2011; Quinlan and Hall 2010). Chromatin interactions in BEDPE format were overlapped with TADs in BED format using the `pairToBed` tool with the following arguments: `-type either` (default parameters), such that a TAD was considered associated with a given interaction if either anchor of the interaction pair overlapped the TAD. In this manner, each set of differential chromatin interactions was assigned its associated TADs, enabling downstream characterization of TAD-level changes linked to specific loop alterations.

High-affinity site integration (Micro-C)

High-affinity sites (HAS) were integrated with the consensus merged sets of differential chromatin interactions using the pybedtools suite (v0.11.0) (Dale et al. 2011; Quinlan and Hall 2010). Chromatin interactions in BEDPE format were overlapped with HAS in BED format using the `pairToBed`

tool with the following arguments: -type either (default parameters), such that a HAS was considered associated with a given interaction if either anchor of the interaction pair overlapped the HAS. In this manner, each set of differential chromatin interactions was assigned its associated HAS, enabling downstream characterization of HAS-level changes linked to specific loop alterations.

Dosage-compensated gene integration (Micro-C)

To identify dosage-compensated genes (DCGs) in the *Drosophila* genome, we used the *Drosophila* S2 chromatin 9-state model (Kharchenko et al. 2010). Specifically, state 5, which corresponds to actively transcribed exons on the male X chromosome undergoing dosage compensation, was used to annotate genes subject to dosage compensation. Using the pybedtools suite (v0.11.0) (Dale et al. 2011; Quinlan and Hall 2010), we overlapped the chromatin state annotations with *Drosophila* gene coordinates using the intersect tool with the -wa and -wb parameters to retain both overlapping regions. All genes overlapping with state 5 were extracted, deduplicated, and designated as DCGs.

DCGs were then integrated with the consensus merged sets of differential chromatin interactions using pybedtools. Chromatin interactions in BEDPE format were overlapped with DCGs in BED format using the pairToBed tool with -type either (default parameters), such that a DCG was considered associated with a given interaction if either anchor of the interaction pair overlapped the gene. Additionally, we performed a second overlap using -type ospan (outer span), which captures genes falling entirely within the genomic span of an interaction rather than overlapping an anchor. This allowed us to differentiate between DCGs contacted by loop anchors and those encompassed within loops on the X chromosome. Together, these approaches enabled comprehensive characterization of DCG-level changes linked to specific loop alterations.

Differential Micro-C chromatin loop annotation using Hi-ChIP chromatin loops (Micro-C, Hi-ChIP)

Differential Micro-C chromatin interactions were annotated using transcription factor (TF)-specific Hi-ChIP chromatin interactions for CLAMP and GAF. Because the Hi-ChIP experiments were performed exclusively under control conditions, only the downregulated Micro-C chromatin interactions in each knockdown condition were annotated. Each set of downregulated Micro-C inter-

actions was overlapped with the corresponding Hi-ChIP dataset using the pairToPair tool (-type both; default parameters) within the pybedtools suite (v0.11.0) (Dale et al. 2011; Quinlan and Hall 2010). The Hi-ChIP interaction sets were generated at 1kb resolution, whereas the Micro-C differential interaction sets spanned 1 kb, 5kb, and 10kb resolutions. To standardize the comparison, we extended the overlap windows for the 1kb and 5kb Micro-C interaction sets using -slop 4500 and -slop 2500, respectively, thereby enabling all overlaps to be performed using uniform 10 kb windows. Downregulated Micro-C interactions that overlapped with Hi-ChIP interactions were classified as TF-specific direct interactions, while those without overlap were classified as indirect interactions, which are interactions lost as a downstream consequence of TF knockdown rather than through direct TF-mediated contact.

Aggregated peak analysis plots (Micro-C, Hi-ChIP)

Micro-C differential chromatin interactions and Hi-ChIP TF-specific chromatin interactions were visualized as aggregate peak analysis (APA) plots using the coolpup.py suite (v0.9.5) (Flyamer et al. 2020). APA plots display the average contact signal centered on loop focal points, where the central pixel corresponds to the bin containing the differential or TF-specific chromatin contact. The resolution of the central bin and the total size of the displayed region, including the flanking sequences, vary by plot and are specified in the corresponding figure legends. The score displayed in the upper-left corner of each APA plot represents the enrichment of the central pixel intensity relative to the background.

Drosophila chromatin state integration (Micro-C)

Micro-C differential chromatin interaction sets were integrated with the *Drosophila* S2 9-state chromatin model (Kharchenko et al. 2010) to annotate loop anchors using the pybedtools suite (v0.11.0) (Dale et al. 2011; Quinlan and Hall 2010). Two distinct approaches were used for this integration.

In the first approach (Fig. 2.1), differential chromatin interactions in BEDPE format were overlapped with chromatin states in BED format using the pairToBed tool with -type either (default parameters), such that a chromatin state was considered associated with a given interaction if either anchor overlapped the state. Under this scheme, a single interaction could appear multiple times if its anchors overlapped with more than one chromatin state. This approach was used to

determine the overall proportion of differential interactions overlapping each genome-wide chromatin state, as well as to characterize loop size distributions for interaction subsets associated with specific chromatin states.

In the second approach (Fig. 2.4), both anchors of each chromatin interaction were independently annotated with their overlapping chromatin states using a custom Python script, generating chromatin state pairs for each interaction. This enabled analysis of the frequency and loop characteristics specific to each pairwise combination of chromatin states. As in the first approach, an interaction could appear multiple times if its anchors mapped to more than one chromatin state pair.

Known and de novo motif discovery

Known and *de novo* motif discovery was performed at loci of interest using the HOMER motif discovery toolkit (v5.1) (Heinz et al. 2010). Specifically, the findMotifsGenome.pl script was run against the *Drosophila* dm6 genome assembly with the following parameters: -size 1000 -len 8,10,12,14,16,20 -mis 4 -S 10, specifying a 1kb window centered on each locus, motif lengths of 8–20bp, up to 4 mismatches, and a maximum of 10 motifs returned. The specific input loci varied by analysis and are detailed in the corresponding figure legends.

roX2-RDD chromatin loop integration

Prior to integrating our Micro-C differential chromatin interaction sets with the *roX2*-RDD chromatin loops (GSE207720) produced by Tian et al. (2024), we normalized anchor bin sizes across all datasets to a uniform 5kb resolution using custom Python scripts. This normalization was necessary because the RDD method generates chromatin loops with variably sized anchors, which would otherwise confound direct comparison with our fixed-resolution Micro-C-derived loops. Although normalization reduced the total number of *roX2*-associated chromatin loops, the remaining set was sufficiently robust for downstream analysis. Complete loop counts for all datasets after normalization, including differential CLAMP- and GAF-dependent loops and *roX2*-RDD loops, are reported in Table SS2.2.

Comprehensive multi-omics visualization

Contact matrices, chromatin loops, bigWig signal tracks, and loci of interest were visualized using the CoolBox API (v0.3.9) (Xu et al. 2021). The following publicly available datasets were incorporated for visualization: H3K36me3 and H4K16ac ChIP-seq data from GSE29537, and MSL2 ChIP-seq data from GSE37864.

2.6 Data Access

Micro-C and Hi-ChIP data is freely accessible via the Gene Expression Omnibus (GSE322508, GSE322509).

2.7 Competing interest statement

The authors declare no conflicting nor competing interests.

2.8 Acknowledgments

We are grateful for the HHMI Gilliam award to J.A., the Blavatnik Family Fellowship to K.C., and the NIH R35 GM153389 award to E.L.

2.9 Author Contributions

J.A. and K.C. performed experiments, computational analysis, and wrote the manuscript. M.A. and L.C.S.A. contributed to computational analysis. M.A., A.A., C.G., and M.R. contributed to cell maintenance, Hi-ChIP experiments, and Micro-C experiments. M.W.S. and K.G. performed Western blot experiments and contributed to methods. E.L. advised on experimental design, analysis, and manuscript writing.

2.10 Supplemental Figures

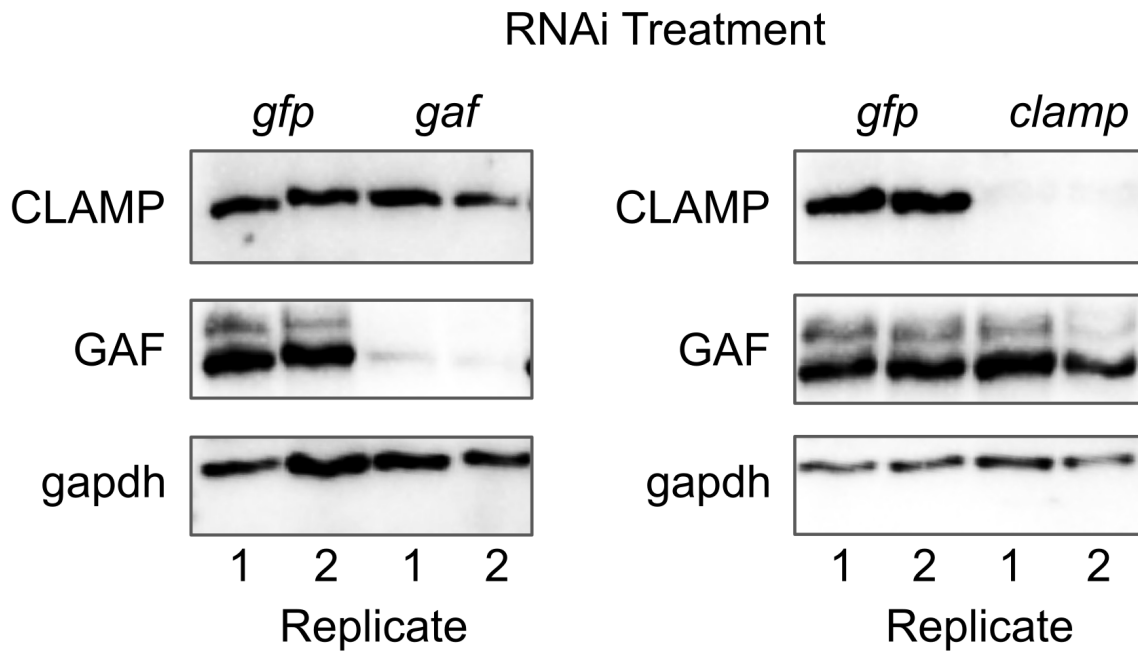


Figure S2.1: **Depletion of CLAMP or GAF in Micro-C samples.** Western blot analysis of protein extracts from cells treated with either *clamp* RNAi (*clampi*) or *gaf* RNAi (*gafi*). Blots were probed with both anti-CLAMP and anti-GAF antibodies across all conditions to confirm knockdown specificity. Treatment with *clampi* results in loss of CLAMP protein and treatment with *gafi* results in loss of GAF protein. anti-GAPDH antibody was used as a loading control to ensure equal amounts of protein lysate was loaded across all lanes.

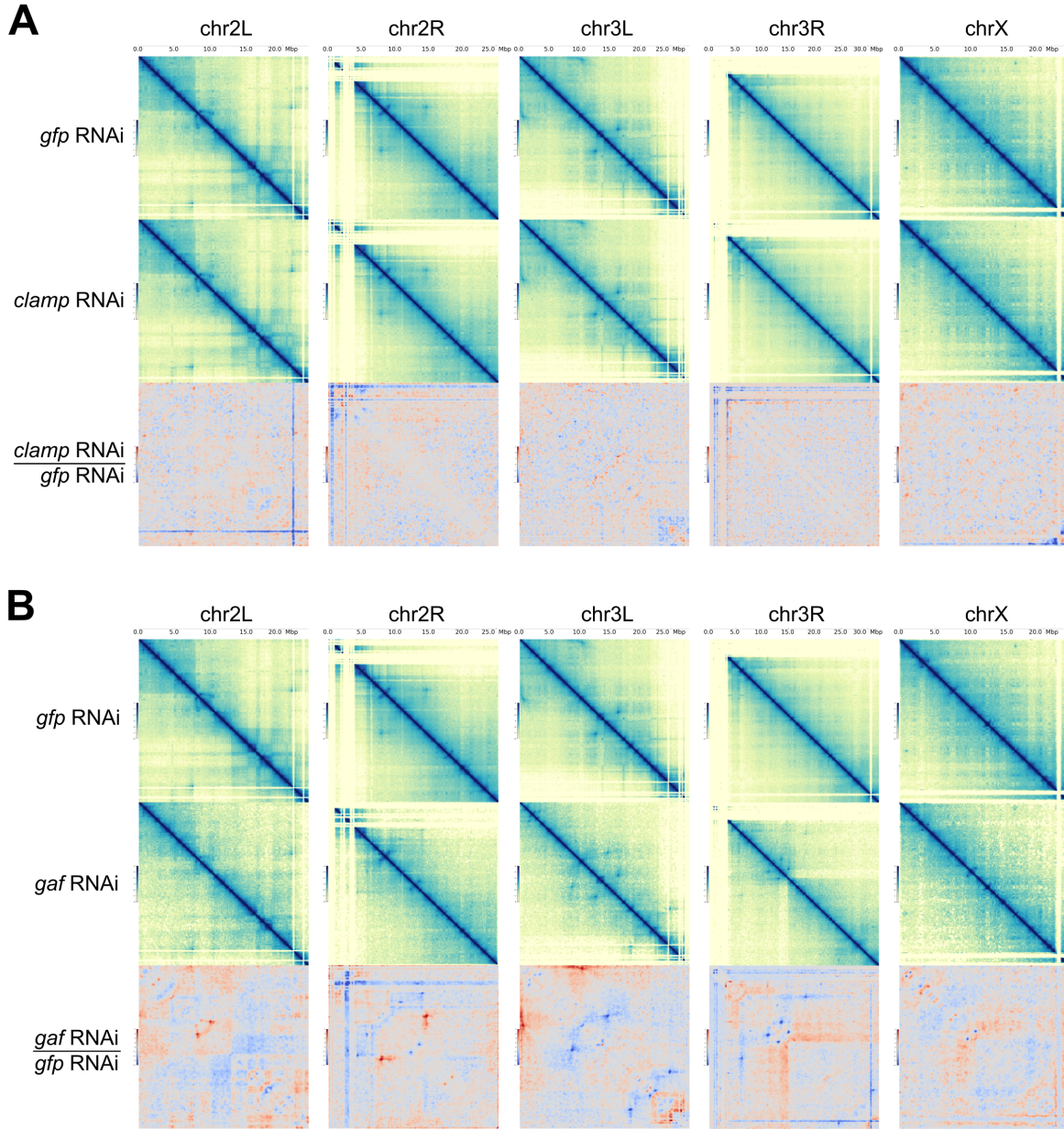


Figure S2.2: **Micro-C contact matrices across all *Drosophila melanogaster* chromosomes.** Micro-C control and post-RNAi treatment contact matrices are shown for all *D. melanogaster* chromosomes: chr2L, chr2R, chr3L, chr3R, and chrX. **(A)** Differential Micro-C matrices for *clamp* RNAi. First row: normalized contact maps for control *gfp* RNAi condition. Second row: normalized contact maps for *clamp* RNAi condition. Third row: differential contact matrices (*clamp* RNAi / *gfp* RNAi) where areas in red indicate increased contact frequency in *clamp* RNAi compared to *gfp* RNAi, while areas in blue indicate decreased contact frequency, with lighter colors representing smaller differences. **(B)** Differential Micro-C matrices for *gaf* RNAi. First row: normalized contact maps for the control *gfp* RNAi condition. Second row: normalized contact maps for *gaf* RNAi condition. Third row: differential contact matrices (*gaf* RNAi / *gfp* RNAi).

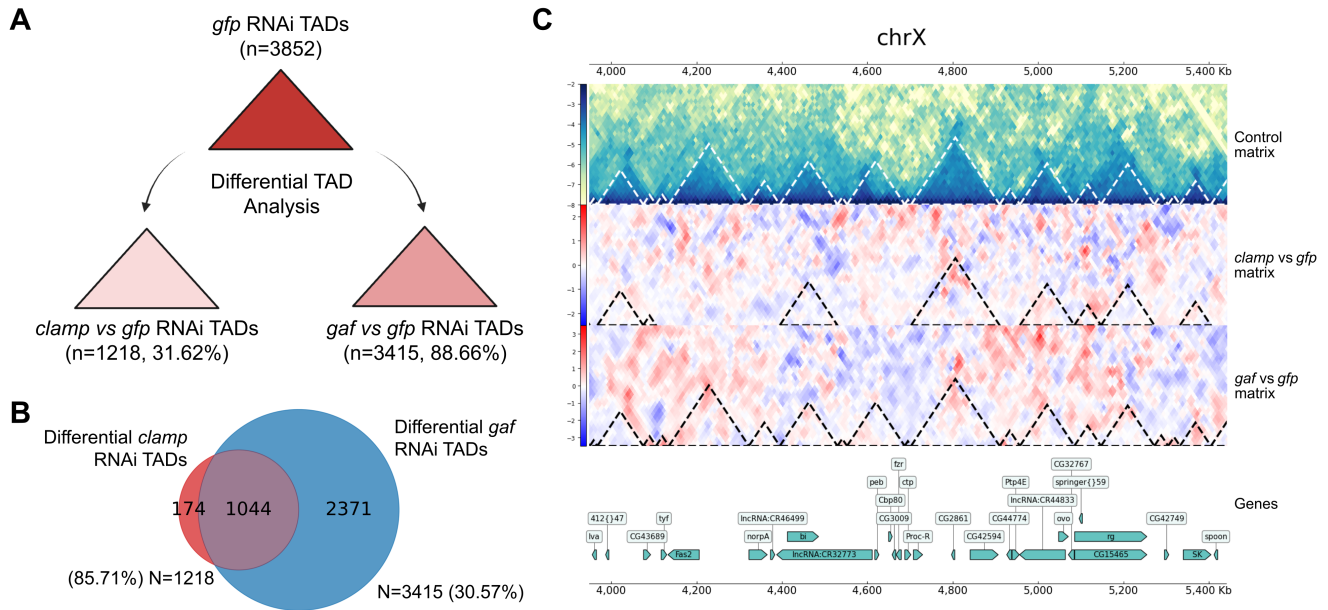


Figure S2.3: TAD organization changes across the genome upon knockdown of CLAMP and GAF. (A) Quantification of *gfp* RNAi TADs and differential TAD comparison after treatment. A total of n=3852 reference TADs were defined from the *gfp* RNAi control data (top triangle). A comparison to *clamp* RNAi (left) reveals n=1218 differential TADs, accounting for 31.62% of all reference TADs. A comparison to *gaf* RNAi (right) reveals n=3415 differential TADs, accounting for a larger 88.66% of reference TADs. This shows a greater genome-wide impact of *gaf* knockdown on TAD structure. (B) Extensive Overlap of CLAMP- and GAF-Dependent TADs. A Venn diagram illustrates that CLAMP and GAF share a large proportion of TAD changes following individual knockdown. Overlap between the sets of differential TADs identified for *clamp* RNAi (n=1218, red) and *gaf* RNAi (n=3415, blue) reveals a significant majority of CLAMP-dependent differential TADs (85.71%, N=1218) are also identified as GAF-dependent differential TADs. This overlap represents a subset (30.57%, N=3415) of the GAF-dependent differential TADs. (C) Differential TADs on the X chromosome after *clamp* and *gaf* RNAi knockdown. Micro-C *gfp* RNAi control matrix: White dashed lines indicate the positions of TAD boundaries with blue indicating greater contact and yellow indicating lower contact frequencies. *Clamp* vs *gfp* matrix: black dashed lines indicate differential TAD boundaries on the X chromosome after *clamp* RNAi treatment. *Gaf* vs *gfp* matrix: black dashed lines indicate differential TAD boundaries on X chromosome after *gaf* RNAi treatment, showing *gaf* knockdown leads to significantly different TADs across the 1,400kb window on X chromosome. Genes: Track of dm6 *Drosophila* X-chromosome genes across the same 1,400kb window.

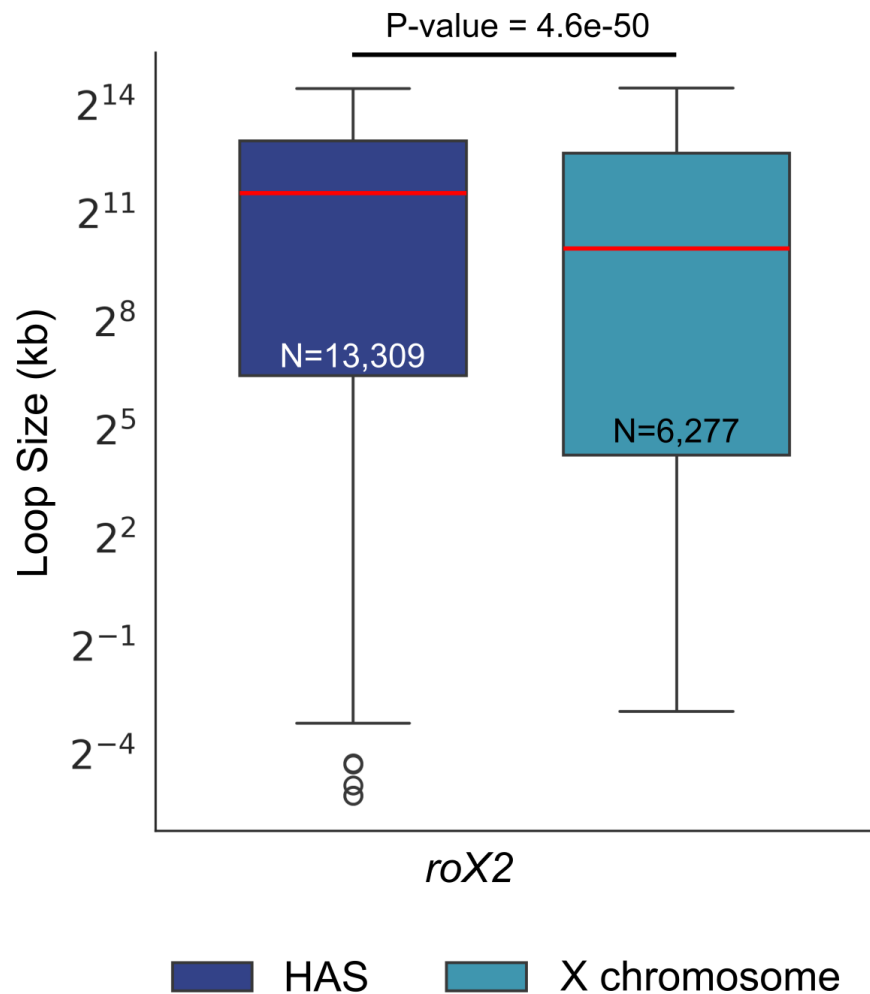


Figure S2.4: **The sizes of *roX2*-associated chromatin loops are significantly larger at HAS than on the X chromosome.** Box plot comparing the distribution of loop sizes (kb) at High Affinity Sites (HAS, dark blue, N=13,309) versus the X chromosome (light blue, N=6,277). Median size (red lines) of *roX2*-associated chromatin loops is significantly larger at HAS than non-HAS loops on X chromosome ($P=4.6 \times 10^{-50}$), as determined by a Mann-Whitney Test.

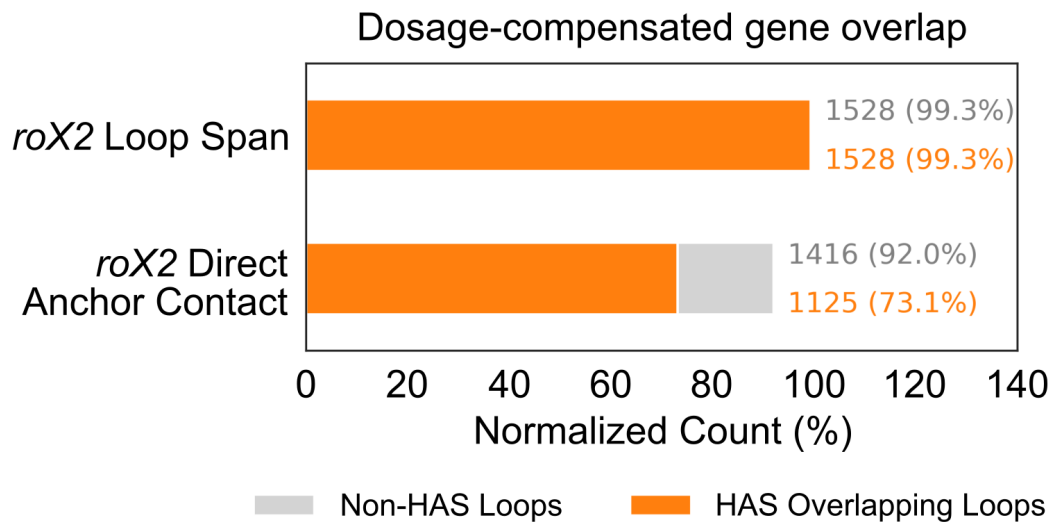


Figure S2.5: **Majority of *roX2*-associated chromatin loops encompassing dosage compensated genes are at HAS on the X chromosome.** Stacked bar plot quantifying the overlap between dosage-compensated genes on the X chromosome and two categories of *roX2*-associated Micro-C loops: HAS Overlapping Loops (orange) and Non-HAS Loops (gray). Assessment of DC genes located within chromatin loop spans associated with *roX2* versus DC genes directly at *roX2*-associated chromatin loop anchors. Nearly all identified dosage-compensated genes in *roX2*-associated chromatin loops (99.3%, n=1528) overlap with HAS. A high proportion (73.1%, n=1125) of dosage-compensated genes are located directly at HAS-overlapping loop anchors, compared to only 18.9% (the remaining portion of the 92.0% total) found at anchors of loops not associated with an HAS.

2.11 Supplemental Tables

Table S2.1: All detected differential Micro-C chromatin interactions.

All Detected Differential Chromatin Loops						
RNAi Comparison	Regulation	Resolution (Kb)	No. of Loops	No. of Loops in Comparison & Regulation Set	Percentage of Comparison Regulation Set	P-value Cutoff
<i>clamp vs gfp</i>	Upregulated	1	8316	34155	24.35%	0.01
<i>clamp vs gfp</i>	Upregulated	5	13539		39.64%	0.01
<i>clamp vs gfp</i>	Upregulated	10	12300		36.01%	0.01
<i>clamp vs gfp</i>	Downregulated	1	11923	16516	72.19%	0.01
<i>clamp vs gfp</i>	Downregulated	5	2572		15.57%	0.01
<i>clamp vs gfp</i>	Downregulated	10	2021		12.24%	0.01
<i>gaf vs gfp</i>	Upregulated	1	12436	26613	46.73%	0.01
<i>gaf vs gfp</i>	Upregulated	5	8128		30.54%	0.01
<i>gaf vs gfp</i>	Upregulated	10	6049		22.73%	0.01
<i>gaf vs gfp</i>	Downregulated	1	2624	5549	47.29%	0.01
<i>gaf vs gfp</i>	Downregulated	5	1579		28.46%	0.01
<i>gaf vs gfp</i>	Downregulated	10	1346		24.26%	0.01

Table S2.2: Normalization of *roX2*, CLAMP-dependent, and GAF-dependent Loops.

Normalization of <i>roX2</i> , CLAMP-dependent, and GAF-dependent Loops				
Normalization Status	Direct ID	Regulation ID	Sample ID	No. of Loops
No	Direct	Downregulated	CLAMP	1,366
No	Direct	Downregulated	GAF	1,679
No	Indirect	Downregulated	CLAMP	15,150
No	Indirect	Downregulated	GAF	3,870
No	-	-	<i>roX2</i>	90,699
Yes	Direct	Downregulated	CLAMP	1,365
Yes	Direct	Downregulated	GAF	1,678
Yes	Indirect	Downregulated	CLAMP	15,149
Yes	Indirect	Downregulated	GAF	3,869
Yes	-	-	<i>roX2</i>	40,951

Chapter 3: The CLAMP GA-binding transcription factor regulates heat stress-induced transcriptional repression

Joseph Aguilera¹, Jingyue Duan², Kaitlyn Cortez¹, Rachel S. Lee³, Angelica Aragon¹, Mukulika Ray⁴, and Erica Larschan¹

Mukulika Ray and I wrote this chapter with editorial feedback from Dr. Erica Larschan.

This chapter has been reviewed and accepted for publication at *eLife*.

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

²College of Agriculture and Life Sciences, Cornell University, Ithaca, NY

³Duke University School of Medicine, Durham, NC

⁴Department of Biology, University of Massachusetts Boston, MA

3.1 Abstract

To survive exposure to heat stress (HS), organisms activate stress response genes and repress constitutive gene expression, thereby preventing the accumulation of potentially toxic RNA and protein products. Although many studies have elucidated the mechanisms that drive HS-induced activation of stress response genes across species, little is known about the mechanisms that repress constitutively expressed genes. Using nascent RNA-sequencing, we identify the first reported transcription factor (TF) that regulates repression of constitutive genes upon heat stress across species and define direct and indirect mechanisms of action by integrating 3D genomic approaches. We demonstrate that the CLAMP (Chromatin-linked adaptor for MSL complex proteins) GA-binding transcription factor (TF) regulates ~75% of the HS-induced repression in *Drosophila*, a well-established model for understanding the mechanisms of HS-induced gene regulation. Using Micro-C, we demonstrate that heat stress induces widespread changes in local 3D chromatin looping, which are significantly associated with HS-induced transcriptional changes. Overall, we identify CLAMP as the first reported TF that induces HS repression, which modulates 3D chromatin looping through both direct mechanisms and indirect mechanisms. Moreover, we present the highest-resolution heat stress 3D genomic dataset available in *Drosophila*, providing a key resource for generating mechanistic insights into how temperature regulates 3D genomic contacts.

3.2 Introduction

Understanding how external stimuli regulate gene expression is essential for reducing the effects of dynamic changes in abiotic factors on all living organisms. One such external stimulus, temperature, plays a vital role in determining viability and fertility, which often decrease as temperatures rise (Richter et al. 2010). Heat stress (HS) rapidly changes transcription of both heat stress genes, which become activated, and constitutive genes, which become repressed (Ray et al. 2019; Mahat et al. 2016). Therefore, understanding the mechanisms that drive changes in transcription upon heat stress will provide critical insight into how transcription rapidly changes in response to stimuli. Because the heat stress response is rapidly inducible, many discoveries of fundamental transcriptional mechanisms have been made using this system, including pausing of RNA Polymerase II,

which occurs at thousands of genes across species (Guertin et al. 2010).

Heat stress induces two types of transcriptional changes: 1) activation of heat stress-induced genes; 2) repression of constitutively active genes (Vihervaara et al. 2018). Best studied in *Drosophila*, decades of research have defined the mechanisms of heat stress-induced gene activation at heat shock-responsive genes, whose promoters contain specific DNA sequences called Heat Shock Elements (HSEs), bound by Heat Shock Factor (HSF) (Guertin et al. 2010; Guertin et al. 2010; Duarte et al. 2016). In *Drosophila*, the GA-binding transcription factor GAF and transcription elongation factor NELF, which regulates RNA Polymerase II pause release, are also involved in heat stress-regulated gene activation (Duarte et al. 2016; Wu et al. 2003; Li et al. 2013). Furthermore, in mammals, NELF functions in HS-regulated transcriptional repression mediated by temperature-dependent phase transition (Rawat et al. 2021). However, the mechanisms that target repression specifically to constitutively active genes remain unknown.

In contrast to activation, much less is known about the specific DNA-binding factors that target constitutive genes for repression after heat stress. Transcriptional repression involves a decrease in promoter-proximal paused RNA Polymerase II at repressed genes (Mahat et al. 2016; Duarte et al. 2016) and cryptic premature termination of transcripts with impaired elongation rate and decreased processivity (Rawat et al. 2021; Cugusi et al. 2022). A report also suggests that heat stress induces the relocalization of architectural proteins from TAD borders to Polycomb binding sites at the interior of TADs, which changes 3D chromatin architecture to potentially regulate repression (Li et al. 2015). However, the role of 3D chromatin architectural changes during the heat stress response still needs to be better understood, as another report suggests that global 3D architectural changes at the TAD and compartment levels do not occur upon heat stress (Ray et al. 2019). Neither of these studies was conducted at a sufficiently high resolution to detect specific local 3D chromatin loop changes. Therefore, the mechanisms by which 3D chromatin changes regulate the heat stress response require further study.

Furthermore, HSF and GAF are involved in heat-stress-induced activation but not heat-stress-induced repression (Duarte et al. 2016). GAF has context-dependent synergistic or antagonistic relationships with another GA-binding transcription factor, CLAMP (Chromatin-linked adaptor for MSL complex proteins) (Kaye et al. 2018; Duan et al. 2021). Although these two transcription factors have similar GA-rich DNA recognition sequences, variation within seemingly similar

cis-elements drives the context-specific targeting of CLAMP to the male X-chromosome and the histone locus, where it inhibits GAF binding (Kaye et al. 2018; Rieder et al. 2017). Additionally, CLAMP promotes long-range 3D chromatin interactions (Jordan and Larschan 2021) and interacts with Negative Elongation Factor (NELF), which functions in heat shock repression (Rawat et al. 2021; Urban et al. 2017). Therefore, we hypothesized that CLAMP regulates heat stress-induced transcriptional repression because it can directly compete for binding with GAF (Kaye et al. 2018), which regulates heat shock activation, and CLAMP interacts with NELF, which regulates repression (Urban et al. 2017).

To test our hypothesis, we used three approaches: 1) nascent RNA-sequencing (SLAM-seq) to define the function of CLAMP in heat stress-induced transcriptional activation and repression; 2) Micro-C to identify CLAMP-dependent and independent genome-wide, high-resolution local changes in chromatin organization after heat stress, and 3) HiChIP to identify CLAMP-bound 3D chromatin loop anchors associated with heat-stress-dependent transcriptional regulation. We found that, unlike GAF, CLAMP regulates heat-stress-induced transcriptional repression at many more genes than activation, with 75% of all repression being CLAMP-dependent. Moreover, the canonical heat stress-activated genes remain activated in the absence of CLAMP.

Furthermore, we demonstrate that heat stress induces global changes in local 3D chromatin organization at chromatin loop anchors. We found that CLAMP inhibits the formation of 3D chromatin loops during heat stress when it is directly bound to loop anchors, but its occupancy at these anchors does not change. In addition, we compared the occupancy of different insulator proteins and chromatin remodelers at CLAMP-dependent HS-repressed genes prior to heat stress. We found that CLAMP directly regulates HS-induced repression through mediating 3D looping when the following factors are enriched near CLAMP binding sites prior to HS: insulator proteins Ibf1/2, ZIPIC, and the chromatin remodeler ISWI, but not Polycomb. In contrast, CLAMP also indirectly regulates repression at loci where it is not bound, potentially by either regulating local 3D chromatin loop formation or regulating transcription independent of 3D looping. Future studies are required to identify which factors regulate the indirect effects on 3D looping and transcription in the absence of CLAMP.

Overall, we identify the first DNA-binding factor that directly regulates repression of the majority of genes that are repressed upon heat stress and define how it functions directly at sites where

it is co-bound with specific co-factors and indirectly regulates repression through modulating 3D contacts. Furthermore, we report a key resource to study HS-induced chromatin 3D structural changes because we generate the highest resolution map to date of heat stress-induced 3D genomic contacts. By investigating the mechanism of HS-induced repression, we enriched our overall understanding of how transcriptional repression is regulated in response to environmental cues in a cellular context.

3.3 Results

3.3.1 CLAMP is required for the repression of transcription in response to heat stress

To understand how the CLAMP GA-binding transcription factor regulates nascent transcription after heat stress, we performed high-throughput sequencing of nascent RNA in *Drosophila* Kc cells using Thiol (SH)-linked alkylation for the metabolic sequencing of RNA (SLAM-seq) (Herzog et al. 2017). SLAM-seq quantifies nascent mRNA by measuring 4SU (4-thiouridine) incorporation as thymine-to-cytosine (T to C) conversions. Cells were incubated in 4SU either at room temperature (RT) for one hour (No heat stress, NHS) or at 37°C for one hour (Heat stress, HS), and then immediately processed to identify which mRNAs incorporated 4SU within the total mRNA pool. Next, we compared the mRNAs incorporating 4SU under HS and NHS conditions to determine which transcripts had increased levels after HS (activated genes) and which transcripts had reduced levels after HS (repressed genes) ($\text{padj} \leq 0.05$ and fold change ≥ 1.5).

Consistent with a prior report (Duarte et al. 2016), we identified a subset of transcripts that were activated ($N = 995$) and a larger subset that was repressed ($N = 3,547$) in control (*gfp* RNAi) cells after heat stress (HS) (Fig. 3.1A). HS-activated transcripts encompass all known heat stress-responsive genes, including those for heat shock chaperone proteins. Much less is known about the mechanisms that mediate HS-induced repression compared to activation. We hypothesized that HS-induced repression is mediated by the CLAMP GA-binding protein TF because it is known to compete with GAF, which regulates heat-stress-induced activation (Duarte et al. 2016; Kaye et al. 2018). Since CLAMP also regulates male X-chromosome dosage compensation in addition to global transcription in both males and females, we decided to examine the function of CLAMP

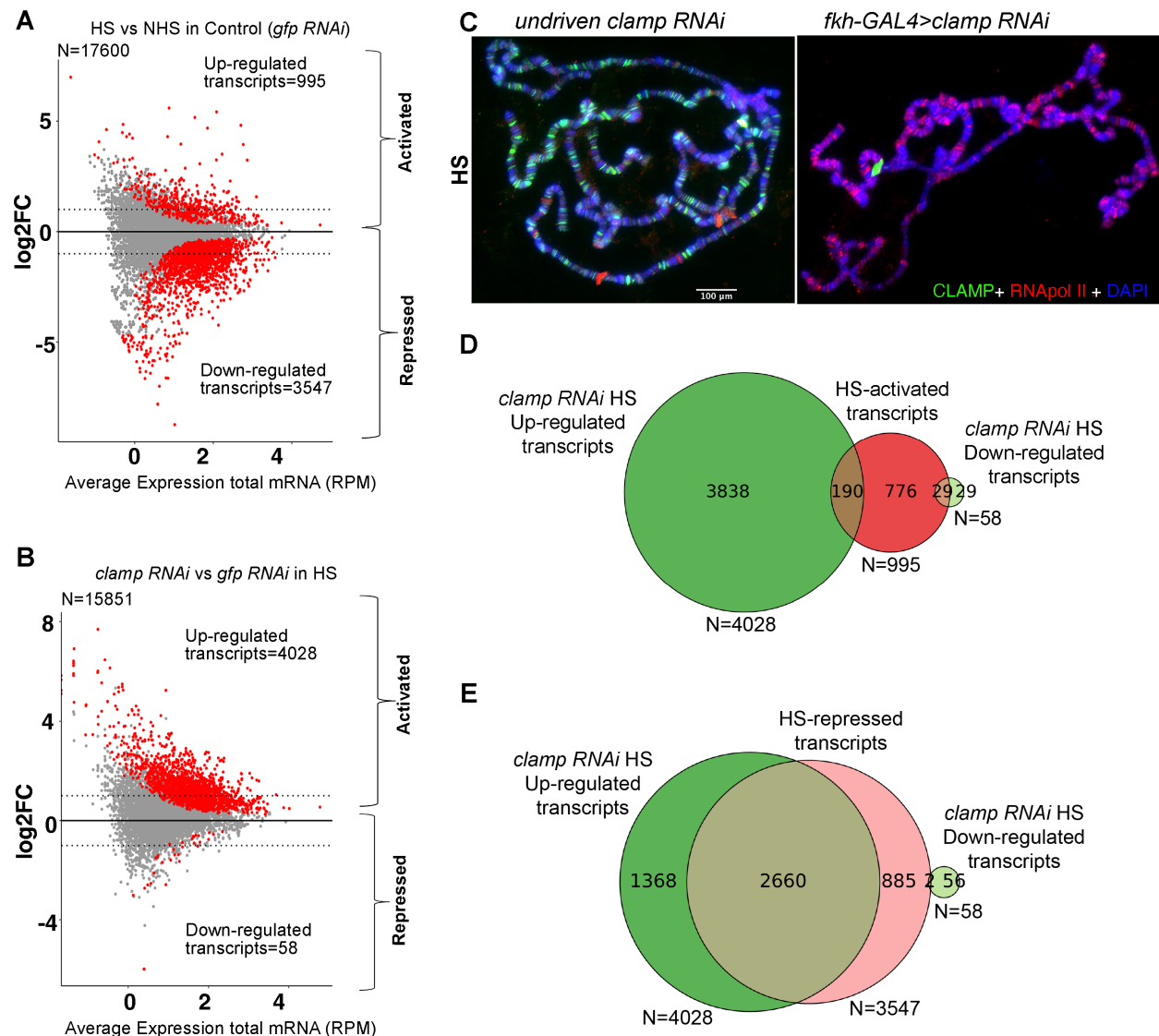


Figure 3.1: CLAMP regulates heat-stress-induced transcriptional repression more strongly than activation (A-B)
MA plots showing significant changes ($FC \geq 1.5$ and $padj \leq 0.05$) in the levels of nascent transcripts (red dots) after HS (heat stress) compared to the control NHS (No heat stress) condition in control (*gfp RNAi*) cells (**A**) and after HS (heat stress) in the absence of CLAMP (*clamp RNAi*) compared to control (*gfp RNAi*) cells (**B**). Dotted horizontal lines in **A** and **B** denote $+1 \log_2 FC$ change cutoffs. (**C**) Immunofluorescence images of HS (37°C for 1 Hr) salivary gland polytene spreads showing the distribution of active RNA polymerase II (red) on chromatin (DAPI: blue) in the presence (undriven *clamp RNAi*) and absence (*fkh-GAL4>clamp RNAi*) of CLAMP (green). The scale bar (100 μ m) applies to both panels in **C**. (**D-E**) Venn diagrams comparing HS-activated transcripts (deep red circle in **D**) and HS-repressed transcripts (light red circle in **E**) from control cells (*gfp RNAi*) to *clamp RNAi* HS Up-regulated (dark green circle) and *clamp RNAi* HS Down-regulated (light green circle) transcripts in *clamp RNAi*-treated cells.

during the heat stress response in the Kc female cell line, which does not undergo dosage compensation.

In the absence of heat stress (NHS), our validated clamp RNAi treatment (Fig. S3.1) (Kaye et al. 2018; Soruco et al. 2013; Urban et al. 2017) causes increased levels of 987 nascent transcripts and decreased levels of 246 nascent transcripts (Fig. S3.2A). Therefore, CLAMP functions as a repressor more frequently than an activator in Kc cells before heat stress (Fig. S3.2A). After HS, unlike control (*gfp* RNAi) cells (Fig. 3.1A), in *clamp* RNAi-treated cells, we observed very little down-regulation of ongoing transcription (Fig. S3.2B). Furthermore, we observed a widespread increase in transcript levels in *clamp* RNAi-treated cells compared with control *gfp* RNAi-treated cells (4,028 upregulated transcripts; 58 downregulated transcripts) (Fig. 3.1B). Therefore, CLAMP functions more frequently as a repressor than an activator of transcription after heat stress in cell culture.

Next, we tested whether CLAMP promotes repression of heat stress response genes *in vivo*. We examined the distribution of active RNA Polymerase II (Pol II) on *Drosophila* larval salivary gland polytene chromosomes after HS (37°C) in control (undriven *clamp* RNAi) and salivary gland driver (*fkh-GAL4>clamp* RNAi) larvae (Fig. 3.1C). Consistent with extensive prior work (Lakhotia et al. 2012; Teves and Henikoff 2011), heat stress restricts the binding of active Pol II (red) to specific cytological positions corresponding to HS-activated loci on control polytene chromosomes (Fig. 3.1C). In contrast, after *clamp* RNAi, active Pol II (red) remains bound genome-wide, including at essential HS-activated loci, suggesting a loss of HS-induced repression but not activation. Notably, *clamp* RNAi flies retain CLAMP at the histone locus, consistent with prior reports (Fig. 3.1C), perhaps because it contains 200 CLAMP binding sites in close proximity to each other (Rieder et al. 2017). Overall, our data show that CLAMP is a stronger regulator of heat stress-induced repression than activation in both Kc cells and flies.

To determine whether CLAMP regulates HS-induced gene activation, we measured the effect of *clamp* RNAi on transcripts that were activated after HS in control cells (Fig. 3.1D). We found that only 2.9% (29 out of 995) of transcripts activated in control *gfp* RNAi cells are dependent on CLAMP (Fig. 3.1D). In contrast, CLAMP inhibits further activation of 19% (190 out of 995) of HS-activated transcripts (Fig. 3.1D). Furthermore, all well-studied canonical HS-induced transcripts, including the *hsp70* gene (Duarte et al. 2016), are still induced even after *clamp* RNAi treatment.

Next, we defined the set of transcripts and genes that require CLAMP for repression after HS (CLAMP-dependent HS repression) by comparing HS-repressed transcripts (N=3547) with transcripts specifically up-regulated in *clamp* RNAi-treated cells (N=4028) and not in control cells (*gfp* RNAi) both under HS conditions (Fig. 3.1E). We found that 75% (2660/3547) of transcripts are CLAMP-dependently repressed after HS (Fig. 3.1E). At the gene level, we identified HS-activated (N=421) and HS-repressed (N=1047) genes (Fig. S3.3A, B), and found that in total 1061 genes are up-regulated and only 30 are down-regulated after HS in the absence of CLAMP (*clamp* RNAi). There is a significant overlap between HS-repressed genes and *clamp* RNAi HS upregulated genes (Fig. S3.3B, N=735, $p=0.00$). All overlaps presented were subjected to Fisher's Exact test combined with FDR correction to determine statistical significance between comparisons (Table S3.1). Thus, the loss of CLAMP has a more substantial effect on HS-induced repression (75% of repressed transcripts are CLAMP-dependent) than on activation (2.9% of activated transcripts are CLAMP-dependent).

3.3.2 Heat stress induces widespread local 3D chromatin changes at chromatin loop anchors

We next asked the question: How does CLAMP regulate transcriptional repression upon HS? Since heat stress induces aggregation of proteins and RNA, especially newly synthesized protein molecules (Wallace et al. 2015), we predict that repression of ongoing transcription ensures that during heat stress, the process of protein production is inhibited at the transcriptional level. Thus, in order to survive under stress by conserving energy and preventing the formation of potentially toxic protein aggregates, the mechanism regulating repression of transcription upon heat shock (HS) needs to be essentially a global phenomenon. Therefore, we hypothesized that HS-induced global changes in transcription repression may be regulated by genome-wide changes, such as global changes in 3D chromatin architecture. Additionally, CLAMP is known to regulate 3D chromatin loop formation and global changes in chromatin organization (Jordan and Larschan 2021), which can be sensitive to changes in temperature, likely through phase transition (Huang et al. 2023), which CLAMP can undergo (Ray et al. 2024).

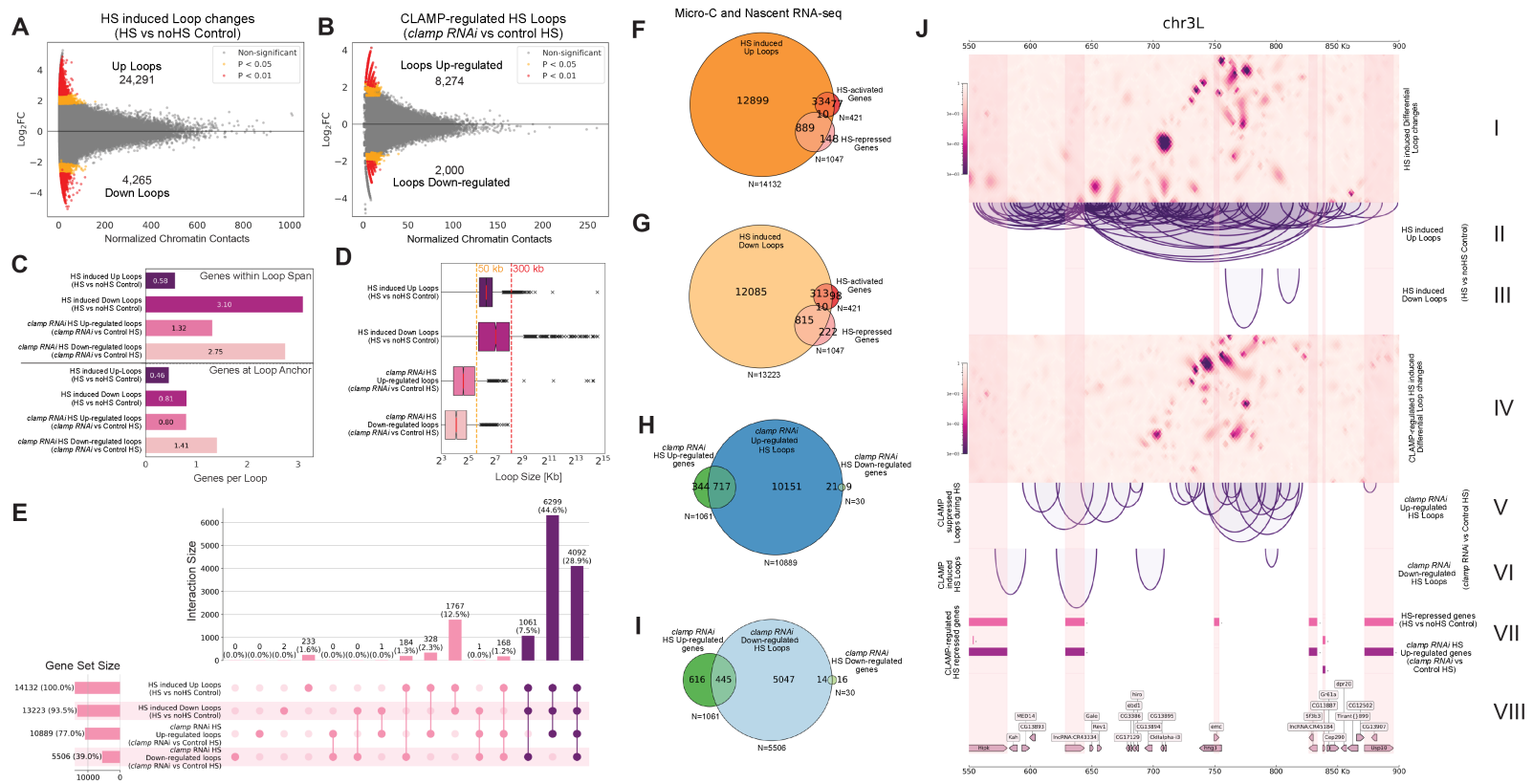


Figure 3.2: Heat stress induces widespread local 3D chromatin changes at the loop anchors (A-B) MA plots for differential loop anchors gained upon HS (HS induced Up loop anchors) or loop anchors lost upon HS (HS induced Down loop anchors) in control cells (A, Up loops=24,291, Down Loops=4,265) and *clamp RNAi* cells (B, Loops Up-regulated=8,274, Loops Down-regulated =2,000), p<0.01. (C) Bar plots show the average number of genes associated with (top box) and present at (bottom box) the HS-induced genome-wide chromatin loop anchors and HS-induced CLAMP-dependent chromatin loop anchors. (D) Box plots showing average loop size of HS-induced genome-wide changing chromatin loops and CLAMP-dependent HS-induced differential chromatin loops with significant difference between average chromatin HS-induced up and down loop sizes, both in case of HS-induced genome-wide changing chromatin loops (Mann–Whitney Test, FDR corrected p=5.63E-186) and CLAMP-dependent HS-induced differential chromatin loops (Mann–Whitney Test, FDR corrected p=3.97E-70), and also between HS-induced genome-wide changing and CLAMP-dependent HS-induced differential chromatin loops themselves (Mann–Whitney Test, FDR corrected p=0). (E) Upset Plot showing genes associated with differential chromatin loop changes and cardinality of every category combination seen as chromatin loop changes after HS in control (top two rows) and *clamp RNAi*-treated (bottom two rows) cells. The first four columns show chromatin loop-associated genes only in each category, and the following six columns show those in each combination of exactly two named sets, the four columns after that combination of three named sets, and the last column shows genes associated with chromatin loop changes in all four categories. HS-associated CLAMP-dependent chromatin loop changes are mostly represented in the last three columns/bars. (F-I) Venn diagrams comparing HS-associated chromatin loop changes (genes within the loop) and HS-activated and repressed genes (F and G); as well as CLAMP-dependent chromatin loop changes during HS with CLAMP-dependent HS-induced transcriptional changes (H and I), showing significant (Fisher's Test) overlap between HS-associated Up loops (F) and Down loops (G) with HS-repressed genes; and CLAMP-dependent HS-associated Down loops (H, *clamp RNAi* up-regulated HS-loops) and Up loops (I, *clamp RNAi* down-regulated HS-loops) with CLAMP-dependent HS-repressed genes (*clamp RNAi* HS up-regulated genes). FDR corrected p-values are 7.487 E-09 (F), 2.53 E-4 (G), 2.0 E-6 (H), and 4.290 E-15 (I), respectively. (J) 350 kb region on 3L showing differential contact interaction results of selfish algorithm on HS and NHS micro-C contact map (I) with associated chromatin loops that are formed (II) and inhibited after HS (III). CLAMP-dependent differential contact interaction results of the selfish algorithm are shown on *clamp RNAi*-treated HS and control HS micro-C contact maps (IV) with CLAMP inhibited (V) and favored (VI) chromatin loops. The color bar in I and IV indicates the q-value (BH adjusted p-value) produced from the DCI analysis. Darker color means this interaction has a lower q-value; that is to say, two contact maps are more diverse at this location. Bed files for differential nascent RNA-seq (SLAM-seq) data from HS and NHS *clamp RNAi*-treated and control cells show the CLAMP-dependent HS-repressed genes (VII) among the corresponding genes within the genomic region associated with the loops (VIII).

The function of 3D chromatin organization in the response to heat shock (HS) has remained unclear because the resolution of prior methods was low, and newer high-resolution techniques were not available when these experiments were conducted. One prior report suggests that HS induces the re-localization of architectural proteins from TAD borders to increase long-range interactions enriched with Polycomb marks, thereby regulating repression (Li et al. 2015). In plants, HS alters the 3D arrangement of promoter-enhancer interactions to regulate the activation of heat shock-responsive genes (Huang et al. 2023; Sun et al. 2020; Liang et al. 2023). In contrast, other reports could not find a correlation between global three-dimensional chromatin changes and transcriptional changes after heat shock (Ray et al. 2019). Unfortunately, prior studies on 3D chromatin and HS used the Hi-C technique with a resolution of ~40 kb at anchor regions, making it challenging to understand the function of specific chromatin-associated factors that bind to regions that are approximately 0.1 kb (Hsieh et al. 2020). However, using the newer micro-C technique to measure 3D chromatin changes after HS, at a resolution of 1-5 kb, could make it possible to define local changes in chromatin organization marked by chromatin loops (Li et al. 2023) that are regulated by a TF like CLAMP.

Thus, using the micro-C technique, we first investigated global changes in 3D chromatin loops at a resolution of 2.5kb after HS in control samples (*gfp* RNAi) (Fig. 3.2A) and then HS induced global changes in chromatin loops regulated by CLAMP (Fig. 3.2B) by performing the micro-C method in *clamp* RNAi-treated cells under HS and NHS conditions using a resolution of 2.5 kb to examine gene-level transcriptional changes (Fig. 3.2C). Moreover, this is the first high-resolution *Drosophila* data set, identifying 3D chromatin changes associated with heat stress that others can mine for 3D chromatin changes associated with HS at specific genomic sites.

First, we identified widespread HS-induced changes in chromatin loops that were increased in frequency (Up loops, N = 24,291) and those decreased in frequency (Down loops, N = 4,265) by HS (Fig. 3.2A, $p < 0.01$). Interestingly, *clamp* RNAi enhanced HS-induced loop formation (*clamp* RNAi HS Up-regulated Loops, N = 8,274) more frequently than it decreased HS-induced loop formation (*clamp* RNAi HS Down-regulated Loops, N = 2,000). Thus, during heat stress, CLAMP primarily functions as an inhibitor of chromatin loop formation (Fig. 3.2B, $p < 0.01$).

Next, we defined how genome-wide local 3D changes in chromatin loop formation correspond to gene distribution at two classes of sites: 1) within the chromatin loops and 2) at the loop anchors

(Fig. 3.2C). This analysis allowed us to determine if HS-associated chromatin loop changes are related to HS-induced transcriptional repression by comparing the chromatin-loop-associated genes with CLAMP-dependent HS-repressed genes. As expected, at the loop anchors which are 2.5 kb in length approximately one gene was associated within both HS-induced Up loops and Down loops independent of *clamp* RNAi, because the average gene size is 3 kb in *Drosophila*. However, when we evaluated the number of genes within the loops, we found significant differences across subcategories with down-regulated loops spanning more genes (3 genes) than upregulated loops (1 gene) independent of *clamp* RNAi (Fig. 3.2C).

Next, we tested whether there were more genes in a loop because the loops were larger in length or because the genes were more densely packed within the same sized loop (Fig. 3.2D, Table S3.1). For HS-induced chromatin loops, the HS-induced Down loops were significantly larger (avg 627 kb) than HS-induced Up loops (avg 95 kb), and thus had more genes within each loop (Fig. 3.2C, D). However, interestingly, the average size of both *clamp* RNAi HS Up-regulated and Down-regulated loops was significantly lower than that of control HS-induced chromatin loops. Therefore, CLAMP regulates the chromatin contacts that form shorter gene-rich loops. Moreover, CLAMP regulates the majority (~85%) of all HS-associated changes in chromatin loops, and most of the time it prevents their formation (Fig. 3.2E).

After determining that CLAMP regulates HS-induced chromatin changes at gene-rich genomic regions, we asked whether these HS-associated 3D chromatin changes were related to HS-induced transcriptional changes (Fig. 3.2F-I). We compared HS-regulated genes with genes located between (Fig. 3.2F-I) and at loop anchors (Fig. S3.4A-D) of the HS-induced chromatin loops. Almost all HS-activated and HS-repressed genes overlapped with HS-induced loop changes (Fig. 3.2F and 3.2G, Table S3.1). This was unsurprising considering we identified whole-genome-wide chromatin loop changes induced by HS. However, we also found that most CLAMP-dependent HS-repressed genes (*clamp* RNAi HS Up-regulated genes) overlapped with *clamp* RNAi HS Up-regulated and Down-regulated loops (Fig. 3.2H and 3.2I, Table S3.1). Interestingly, more (N=717, 68%) of the CLAMP-dependent HS-repressed genes overlapped with HS loops suppressed by CLAMP (*clamp* RNAi HS Up-regulated loops) compared with HS loops promoted by CLAMP (N=445, 42%) (Fig. 3.2H-I). Overall, these data are consistent with a function for CLAMP in repressing heat shock-induced gene expression by reducing the formation of chromatin loops that normally activate

genes.

To visualize the association of CLAMP-mediated chromatin loop changes with CLAMP-dependent HS-repression, we identified a 350 kb gene-dense region on chromosome 3L that is associated with a large number of HS-induced chromatin loops and multiple HS-repressed genes (Fig. 3.2J). CLAMP regulates a subset of the chromatin loops in this region which are associated with CLAMP-dependent HS-repressed genes. Thus, we demonstrate that CLAMP regulates 3D chromatin changes during heat shock (HS) which are associated with CLAMP-dependent HS-induced transcriptional repression.

3.3.3 CLAMP inhibits chromatin loop formation during HS and regulates repression of associated genes through both direct and indirect mechanisms

The next question we asked was whether CLAMP has a direct effect on 3D chromatin changes and related transcriptional changes after HS or if it functions through a combination of direct effects at CLAMP-bound loci and indirect mechanisms as is common for most TFs. High-resolution chromatin organization data in HS *clamp* RNAi cells reveal that CLAMP regulates chromatin loop rearrangement during HS (Fig. 3.2B) at loops associated with CLAMP-dependent repressed genes (Fig. 3.2I, J). However, to understand the direct role of CLAMP in transcriptional regulation after HS at high resolution in 3D, we used HiChIP to define which changes were directly associated with CLAMP binding to loop anchors. HiChIP measures TF binding and 3D loop anchors simultaneously at less than 1kb resolution (Bhattacharyya et al. 2019; Gutierrez-Perez et al. 2019; Mumbach et al. 2016), making it possible to define both TF binding and associated 3D chromatin loops. Therefore, we used HiChIP to investigate whether changes in 3D loop anchor contacts are associated with alterations in CLAMP occupancy on chromatin and HS-induced transcriptional repression.

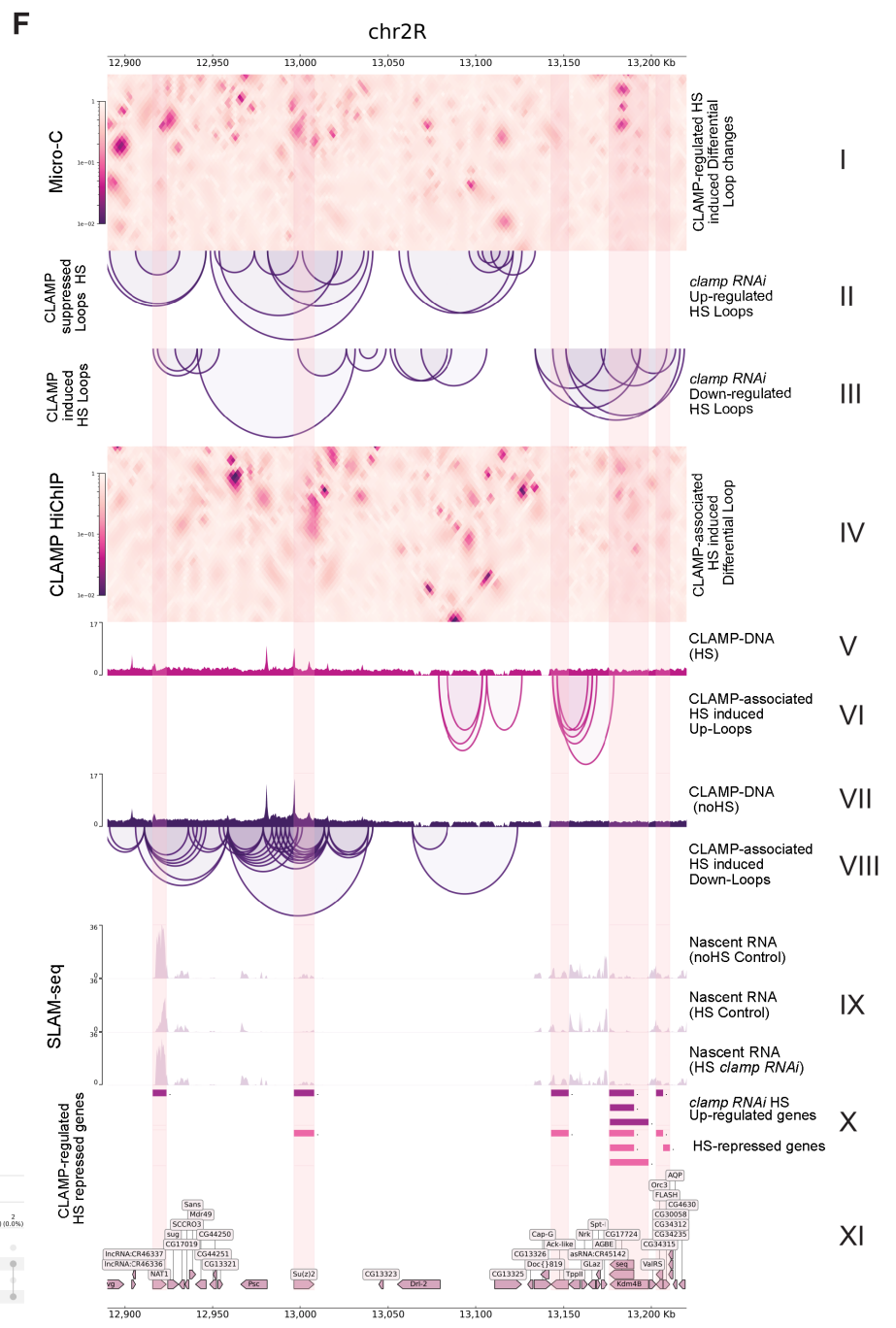
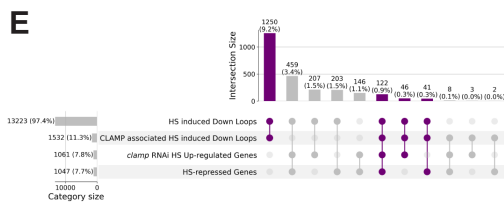
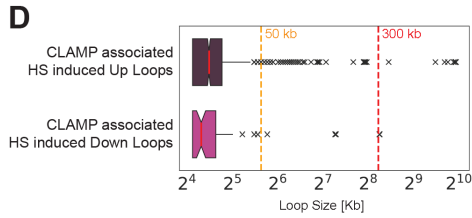
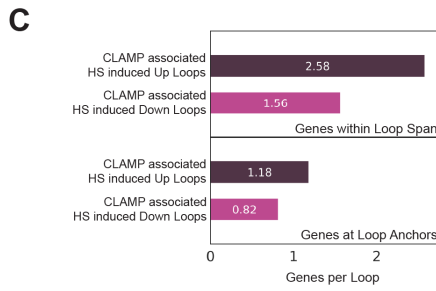
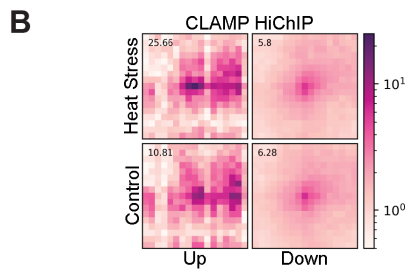
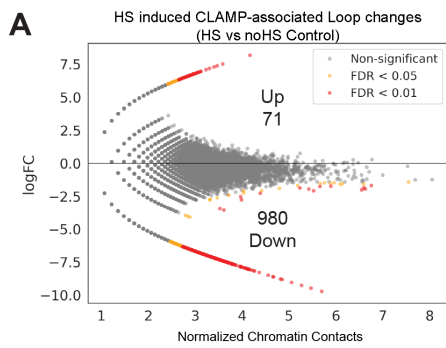


Figure 3.3: **CLAMP regulates chromatin loop changes during HS, directly as well as indirectly, to repress gene transcription** (A) 2D metaplot of CLAMP HiChIP data centered on significant interactions identified from CLAMP HiChIP under NHS or HS conditions for differential loop anchors gained upon HS (HS Up loop anchors, N=71) or loop anchors lost upon HS (HS Down loop anchors, N=980). The score indicates the enrichment of the center pixel compared with the top left corner. (B) CLAMP-bound loop pile-ups for HiChIP data from heat stress (HS) and control cells showing loss and gain of chromatin structures marked by chromatin loops defined in HS and control cells. The color is average interaction frequency across the entire dataset surrounding the defined HS Up and Down loops. A high signal indicates strong interaction between loop anchors. ± 20 kb padding was applied to the central 2.5 kb bin, displaying a 42.5 kb aggregated region. (C) Barplots showing average number of genes associated within (top box) and present at (bottom box) the CLAMP-bound chromatin loop anchors up and down-regulated after Heat shock (HS). (D) Box plots showing significant difference (Mann–Whitney Test, FDR corrected $p=0.0354$) between average loop size of CLAMP-bound chromatin loop anchors up and down-regulated after Heat shock (HS). (E) Upset Plot comparing all (Micro-C data) and CLAMP-bound (HiChIP data) chromatin loops lost after HS with HS-repressed and CLAMP-dependent HS-repressed genes, showing significant (Fisher's Test) overlap between CLAMP-bound HS-associated Down-loops with global genome-wide loss of HS-associated loops ($p=1.26E-115$), HS-repressed genes ($p=9.59E-18$), and CLAMP-dependent HS-repressed genes (*clamp* RNAi HS up-regulated genes, $p=2.57E-19$). (F) 300 kb region on 2R showing CLAMP-regulated differential contact interaction results of selfish algorithm on *clamp* RNAi-treated and control HS micro-C contact map (I) with associated CLAMP suppressed (II) and induced (III) chromatin loops after HS. CLAMP-associated differential contact interaction results of the selfish algorithm are shown on HS and NHS CLAMP HiChIP contact maps (IV) with CLAMP-bound (V and VII) directly associated Up (VI) and Down (VIII) chromatin loops. The color bar in I and IV indicates the q-value (BH adjusted p-value) produced from the DCI analysis. A darker color means this interaction has a lower q-value; that is to say, two contact maps are more diverse at this location. Normalized reads and bed files for differential nascent RNA-seq (SLAM-seq) data from HS and NHS *clamp* RNAi-treated and control cells show the CLAMP-dependent HS-repressed genes (IX and X) among the corresponding genes within the genomic region associated with the loops (XI).

We performed HiChIP using validated antibodies (Rieder et al. 2017; Jordan and Larschan 2021; Urban et al. 2017) targeting the CLAMP protein to identify CLAMP-associated loop anchors in Kc cells exposed to HS and matched control NHS Kc cells. Loop anchors bring distal regulatory elements closer to each other, which can regulate activation or repression of gene expression depending on the factors bound (Gutierrez-Perez et al. 2019; Ogiyama et al. 2018). Using the FitHiChIP platform (Bhattacharyya et al. 2019), we first identified the following categories of CLAMP-bound chromatin loops (Fig. S3.5): 1) present in control cells without HS (NHS) (N=1515); 2) shared loops that are present in both control (NHS) cells and HS cells (N=421); 3) loops present only after HS and not in NHS conditions (N=563). We then classified loops into sub-categories based on whether they increased or decreased after heat shock: 1) CLAMP-bound HS induced Up loops gained, and 2) CLAMP-bound HS induced Down loops lost upon HS (Fig. 3.3A, B). We found that most CLAMP-bound chromatin loops are lost (N=980, Down loops) upon HS and very few are gained (N=71, Up Loops), clearly showing that CLAMP-bound chromatin loops are mainly down-regulated after HS. Our analysis shows that of the 980 CLAMP-associated HS induced down loops (HiChIP), 931 loops (95%) only had one anchor associated with CLAMP direct binding, while 49 loops (5%) had both anchors associated with CLAMP direct binding. Of the 71 CLAMP-associated HS induced up loops (HiChIP), all 71 loops (100%) only had one anchor

associated with CLAMP direct binding. Overall, we conclude that the direct effect of CLAMP is to promote the loss of 3D chromatin loops at specific genomic locations after HS more frequently than it promotes the formation of new loops.

Next, we asked whether CLAMP occupancy on chromatin changes when 3D loop contacts change. Therefore, we categorized CLAMP-bound loop anchors lost after HS (HS Down) and gained after HS (HS Up) into the following sub-classes depending on the change in occupancy of CLAMP on chromatin (similar to ChIP-seq) which HiChIP also quantifies in parallel with 3D loop anchors (Fig. S3.6): 1) High Difference (HD) when CLAMP occupancy between HS and NHS conditions is differential by EdgeR (Robinson et al. 2010; Chen et al. 2016); 2) No Difference (ND) when CLAMP occupancy between HS and NHS conditions is non-differential by EdgeR (Robinson et al. 2010; Chen et al. 2016) but exhibits a $<25\%$ difference in ChIP-seq coverage; 3) Low Difference (LD) when CLAMP occupancy between HS and NHS conditions is non-differential by EdgeR (Robinson et al. 2010; Chen et al. 2016) but exhibits a $\geq 25\%$ difference in ChIP-seq coverage. Because each loop has two ends, the loop anchors were further categorized into five types: HD-HD, HD-LD/ND, ND-ND, LD-ND, and LD-LD (Fig. S3.6A-F). We found that most CLAMP-bound chromatin loop anchors gained (Up) (65/71: 92%) and lost (Down) loop anchors (849/980: 87%) upon HS belong to the ND or LD categories, in which CLAMP occupancy on chromatin is not dramatically changed. Thus, we conclude that upon HS, most CLAMP-bound chromatin changes occur through modulation of 3D chromatin contacts between loop anchors rather than through the gain or loss of CLAMP occupancy at specific genomic locations, consistent with the rapid time scale of heat stress-induced transcriptional and phase changes.

We also measured the average number of genes present: 1) at the loop anchors and 2) within the CLAMP-bound HS-induced differential chromatin loops. We quantified the average loop length of CLAMP-bound loops. We found that CLAMP-bound loops were gene-rich, both at the loop anchors and within the loops, with the average CLAMP-bound HS-induced chromatin loop size smaller (avg 38 kb) compared to global HS-induced differential loop size (avg 172 kb) (Fig. 3.3C-D and Fig. 3.2C-D).

Since the majority of CLAMP-bound loops were down-regulated by heat shock, we investigated what proportion of all HS-induced Down-loops were CLAMP-bound chromatin loops and whether they overlapped with HS-repressed genes (Fig. 3.3E). Although most 3D loop changes

after HS are CLAMP-dependent (~85%, Fig. 3.2E), only ~11% of HS-induced Down-loops were CLAMP-bound chromatin loops (Fig. 3.3E), of which 122 sites overlapped with CLAMP-dependent HS-repressed genes. This suggests that CLAMP modulates 3D looping through both direct mechanisms that likely involve specific combinations of co-factors and indirect mechanisms that may involve modulation of the activity of other chromatin loop regulators in the cell.

To visualize the types of loops that CLAMP mediates directly to regulate heat shock repression, we focused on a genomic region on chromosome 2R, where we identified a subset of global 3D chromatin loop changes that are directly bound and regulated by CLAMP (Fig. 3.3F). We show that a subset of CLAMP-bound chromatin loops are differentially affected by HS, and corresponding genes within these loops are repressed after HS. At the CLAMP-associated loops, before and after HS, there is no significant difference in CLAMP occupancy (Fig. 3.3G, V and VII); however, there is differential chromatin loop formation, with a subset of chromatin loops lost while others are formed after HS (Fig. 3.3G, VI and VIII). The repression of these target genes is CLAMP-dependent, as shown by the nascent RNA tracks from HS and NHS control and clamp RNAi-treated cells (Fig. 3.3G, IX, X). Therefore, CLAMP directly represses a subset of its target genes through modulating chromatin looping.

Next, we integrated the CLAMP HiChIP data with the nascent RNA sequencing and micro-C datasets to identify CLAMP-dependent HS-repressed genes (N=735), which are: 1) directly CLAMP-bound and CLAMP-dependent HS-induced Down-loops (14.4%, N=106); 2) indirectly CLAMP-dependent HS-induced Down-loops (52.8%, N=388), and 3) independent of HS-induced loop changes (29.5%, N=217) (Fig. 3.4A). The first gene category comprises repressed genes directly regulated by CLAMP at the 3D chromatin organizational and transcriptional level after HS. In contrast, the second gene group is indirectly regulated by CLAMP at the 3D chromatin organizational level after HS. In contrast, the last group is regulated by CLAMP at only the transcriptional level and not the 3D chromatin organizational level (Fig. 3.4A). Thus, repression of genes after HS due to changes in 3D chromatin architecture at ~67% of HS-repressed genes is either directly or indirectly regulated by CLAMP. In contrast, ~30% of the CLAMP-dependent HS-repressed genes are not associated with concurrent changes in 3D loop structure after HS, indicating a 3D loop-independent transcriptional regulation of a subset of CLAMP-dependent HS-repressed genes (Fig. 3.4A). Therefore, we conclude that CLAMP directly and indirectly inhibits 3D chromatin loop formation during HS at a

subset of gene-rich genomic locations, resulting in their transcriptional repression.

3.3.4 Specific combinations of co-factors present prior to heat stress distinguish the direct from the indirect functions of CLAMP in 3D looping and heat stress repression

Next, we asked the question: What combination of chromatin-associated factors distinguishes CLAMP-dependent repressed genes that are directly versus indirectly regulated by CLAMP? To address this question, we screened publicly available architectural protein datasets (GSE11804, GSE30740, GSE36374, GSE36393, GSE54529, GSE63518, GSE80702, GSE89244) (Li et al. 2015; Gutierrez-Perez et al. 2019; Wood et al. 2011; Yang et al. 2013; Kellner et al. 2012; Van Bortle et al. 2012; Van Bortle et al. 2014; Cubenas-Potts et al. 2017; Rowley et al. 2017), which were previously used to examine co-factors for Pipsqueak (also, a GA-binding transcription factor similar to CLAMP) and Polycomb in 3D chromatin organization (Gutierrez-Perez et al. 2019). We measured the occupancy of known chromatin marks and binding factors prior to HS at the CLAMP-dependent HS-repressed genes (N=735), which are associated with: 1) direct CLAMP-dependent HS-induced Down-loops (N=106); 2) indirect CLAMP-dependent HS-induced Down-loops (N=388); and 3) HS-induced loop change independent (N=217) (Fig. 3.4A, B).

We identified factors present before HS at the CLAMP-dependent genes associated with CLAMP-bound chromatin loops (direct regulation) and those associated with CLAMP-independent chromatin loops (indirect regulation) to identify candidate co-factors involved in directly regulating HS-induced transcription repression by CLAMP. We used a set of random genes not repressed after HS as a control (Fig. 3.4B). Genes in all three categories lacked repressive factors such as Su(Hw) and Pc (Gutierrez-Perez et al. 2019; Kahn et al. 2023) (Fig. 3.4B). Compared with genes that are not repressed by CLAMP, the CLAMP-dependent HS-repressed genes (N=735) (both direct and indirect effects) were enriched for: 1) ISWI, a chromatin remodeler (Dahn et al. 2017) at their TSS and TES (transcription start and end) sites which is known to require CLAMP for its recruitment (Urban et al. 2017); 2) CP190, a BTB-ZnF factor present in most insulator complexes; and 3) ZIPIIC (Zinc-finger protein interacting with CP190), a promoter-bound TF with architectural and insulator function (Cuartero et al. 2014; Maksimenko et al. 2015; Zolotarev et al. 2016; Kaushal et al. 2022)

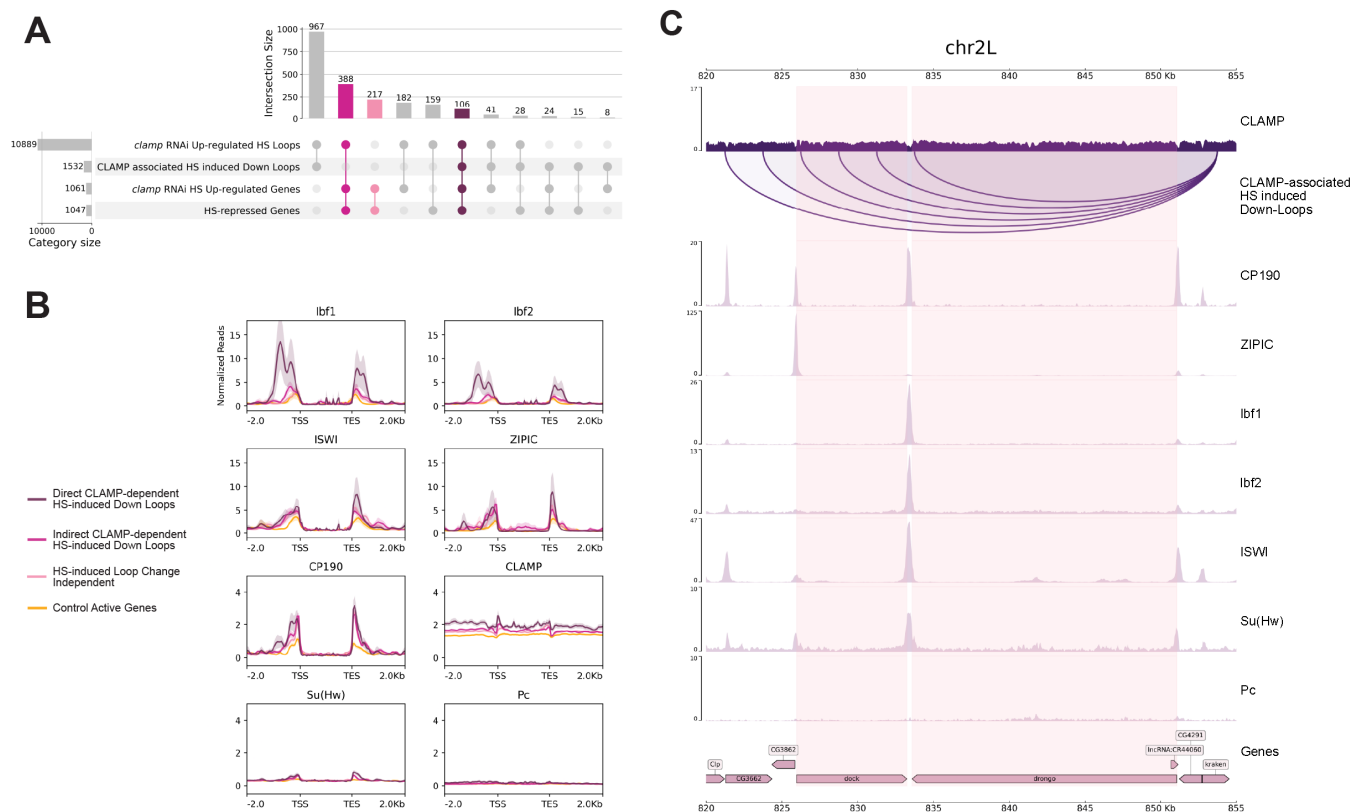


Figure 3.4: HS-associated chromatin loops directly and indirectly regulated by CLAMP have different chromatin-bound factors under NHS conditions (A) Upset plot comparing all CLAMP-dependent (Micro-C data, *clamp* RNAi up-regulated HS Loops) and CLAMP-bound (HiChIP data) chromatin loops lost after HS with CLAMP-dependent HS-repressed genes, showing a significant (Fisher's Test) number of CLAMP-dependent HS-repressed genes are associated with CLAMP-dependent loss of chromatin loops after HS ($p=1.55E-04$). Notably, a significant proportion of these are directly CLAMP-bound ($p=8.78E-32$). (B) Schematic showing three categories of CLAMP-dependent HS-repressed genes based on 3D chromatin organisational changes associated with HS. (C) Average profile for chromatin-bound factors across CLAMP-dependent HS-repressed genes present at direct and indirect CLAMP-dependent HS-induced Down-loop anchors, genes not associated with any loop anchors, and a random set of genes (negative control). (D) 35 kb region on 2L showing distribution of chromatin bound factors at CLAMP-dependent HS-repressed genes dock and drongo and associated CLAMP-bound chromatin loops lost after HS.

(Fig. 3.4B). Interestingly, genes where CLAMP has a direct effect on repression through regulating 3D loops (N=106) were more enriched for the CP190-interacting insulator binding factors: Ibf1 and Ibf2 (Cuartero et al. 2014) (Fig. 3.4B). Thus, we hypothesize that Ibf1 and Ibf2 are important co-factors that regulate the direct function of CLAMP in repression of genes through 3D looping.

The presence of a chromatin remodeler like ISWI and insulator proteins like ZIPIC and CP190 prior to HS at genes that will be repressed may prime these genes for subsequent transcription repression. Upon HS, CLAMP is directly required for the loss of the chromatin 3D loop structure and repression at a subset of these genes that are enriched for Ibf1/2 (Fig. 3.4C). In contrast, genes enriched for ZIPIC and CP190, but lacking Ibf1/2 and CLAMP at the chromatin loops, still undergo 3D chromatin loop loss and repression after HS, but CLAMP is indirectly responsible for their 3D chromatin loop changes. Future work is required to identify the factors that regulate the CLAMP-dependent indirect changes in 3D chromatin organization and transcriptional repression: ZIPIC and CP190 are strong candidate factors to regulate indirect effects. Also, CLAMP-dependent repressed genes (Fig. 3.4B) that are not associated with any 3D chromatin changes do not show enrichment for ISWI or insulator proteins compared to genomic regions undergoing 3D changes. How CLAMP regulates repression independent of 3D looping requires further investigation and may involve modulating Pol II pausing because we have shown that CLAMP directly interacts with the NELF pausing regulator (Urban et al. 2017).

Overall, our studies have identified CLAMP as the first reported DNA-binding factor that regulates HS-dependent repression and identified a mechanism that involves changes in 3D chromatin loops. Moreover, we provide the highest resolution data set for studying 3D chromatin loops after heat stress in *Drosophila*. Our data are consistent with a model in which CLAMP-regulated inhibition of chromatin loop formation after HS is responsible for repression of genes after HS, suggesting that these loops would normally be important for maintaining active transcription (Fig. 3.5). We also identify Ibf1/2 as candidate co-factors that are enriched when CLAMP functions directly versus indirectly in regulating transcriptional repression through 3D looping. Future work will be required to define the combinatorial context-specific interactions between CLAMP, chromatin remodelers, and insulators that determine how genes become repressed upon heat stress.

3.4 Discussion

Temperature is one important abiotic regulator of gene expression that substantially regulates biological functions. Defining how increasing temperature affects gene transcription has not only identified new transcriptional mechanisms but is vital for managing the effects of global changes in temperature on living organisms, including humans, plants, and other animals.

To survive under heat stress, cells must repress the ongoing transcription of constitutive gene transcripts to avoid accumulating toxic protein products, which would lead to cell death. Compared with decades of work on transcriptional activation upon heat stress (Mahat et al. 2016; Guertin et al. 2010; Vihervaara et al. 2018; Duarte et al. 2016), very few reports have provided insight into how HS-induced transcriptional repression is regulated (Mahat et al. 2016; Duarte et al. 2016; Rawat et al. 2021; Cugusi et al. 2022; Li et al. 2015; Goenka et al. 2016). In particular, sequence-specific DNA-binding TFs that target genes for transcriptional repression remained unknown.

Here, we demonstrate that the GA-binding transcription factor CLAMP is the first-reported TF that controls the repression of the majority of genes (Fig. 3.1) and genome-wide rearrangement of 3D chromatin loops (Fig. 3.2E) upon heat stress. In contrast, CLAMP does not regulate any of the canonical genes activated upon heat stress, such as the heat shock protein genes (Fig. 3.1). A significant number of CLAMP-dependent HS-repressed genes overlap with CLAMP-dependent 3D loop changes (Fig. 3.2H-I, Table S3.1), indicating the role of CLAMP in regulating HS-dependent transcription repression via 3D chromatin structure changes. However, the loss of CLAMP-associated loop anchors alone is insufficient to determine whether a gene will be activated or repressed because CLAMP-associated loop anchors are lost upon HS at both repressed and activated genes (Fig. 3.2).

Furthermore, we found that only a subset of CLAMP-dependent repressed genes are associated with the CLAMP-bound 3D chromatin loops that are lost upon heat stress (Fig. 3.4A), suggesting a more restricted direct and a broader indirect role for CLAMP in regulating local changes in 3D genome organization for HS-induced transcription repression. However, CLAMP is the first report of any DNA-binding factor involved in HS-induced changes in gene repression and chromatin organization across species. Also, direct mechanisms of transcriptional dysregulation are always a subset of all transcriptional changes caused by the loss of a transcription factor *in vivo*,

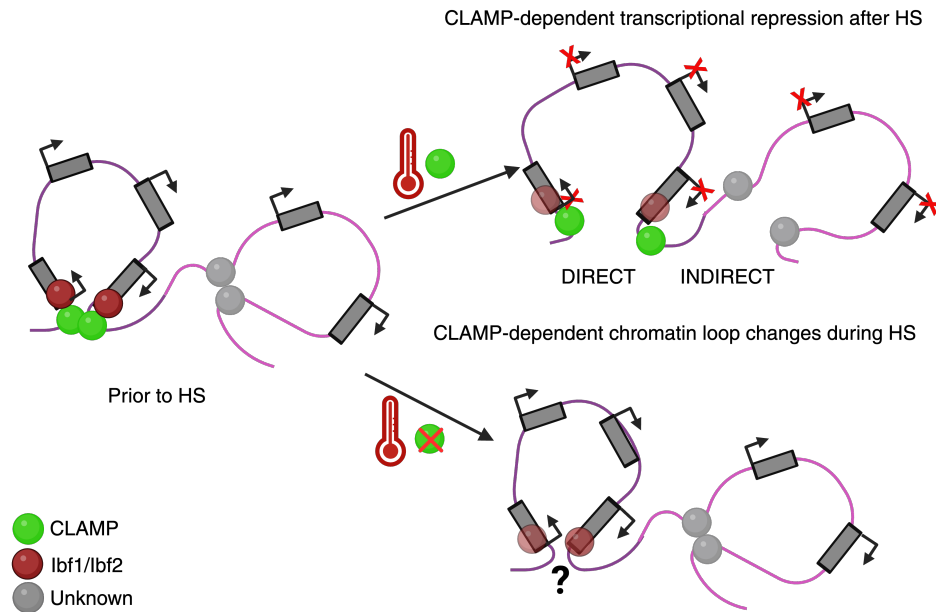


Figure 3.5: **Model for CLAMP-dependent HS-induced transcriptional gene repression**

and therefore we asked whether there are specific co-factors enriched prior to heat stress at genes where CLAMP functions directly compared with sites where it functions indirectly.

On average, CLAMP-dependent HS-repressed genes are modestly enriched prior to heat stress for ZIPIC, a zinc finger protein (Fig. 3.4B). ZIPIC is an architectural protein with homodimerization domains that promotes CP190-mediated insulator activity and boundary formation, which can prevent transcription of nearby genes (Maksimenko et al. 2015; Zolotarev et al. 2016; Kaushal et al. 2022). CP190 is another protein modestly enriched at the CLAMP-dependent HS-repressed genes. CLAMP physically interacts with CP190 and both promote gypsy chromatin insulator activity (Bag et al. 2019). Additionally, the enrichment of ISWI, an essential chromatin remodeler that can be recruited by CLAMP (Urban et al. 2017) at HS-repressed genes compared to control active genes, further supports the hypothesis that a change in chromatin architecture drives HS-induced gene repression. Also, most (67%) CLAMP-dependent HS-repressed genes are associated with HS-dependent chromatin loop changes (N=388+106) regulated by CLAMP, either directly or indirectly (Fig. 3.4A, Table S3.1).

We found that CLAMP-dependent HS-repressed genes, at which CLAMP directly regulates chromatin looping, were enriched for the Ibf1/2 insulator proteins that interact with CP190 (Cuartero et al. 2014) (Fig. 3.5). However, future work is necessary to define the CLAMP-interacting factors at loci where CLAMP indirectly regulates 3D chromatin changes during HS without binding to the

loop anchors, as well as at the smaller group of loci where HS-induced transcriptional repression is loop-independent. Furthermore, how the combinatorial context-specific interactions between all the factors we identified and other unknown factors determine whether a gene will become repressed upon heat stress still needs to be understood.

Also, we have not yet determined how HS results in 3D chromatin changes at CLAMP-bound loops, especially in places where there is no concurrent loss in CLAMP occupancy after HS. We predict that loss of the loop after HS would result in transcriptional repression due to loss of interactions that activate transcription through 3D looping or 2D mechanisms such as RNA Polymerase II pause release (Duarte et al. 2016). In addition, we show that the direct effect of CLAMP on chromatin looping to mediate repression is associated with binding prior to HS of the insulator chromatin factors (Ibf1/2) and not the Polycomb and Su(Hw) complexes. However, it is possible that after heat stress, the occupancy of these factors changes, and the biomolecular mechanisms by which CLAMP and Ibf 1/2 drive 3D chromatin structural changes remain unclear.

CLAMP is an intrinsically disordered protein with a prion-like domain, PrLD (Kaye et al. 2018; Ray et al. 2024). Therefore, we speculate that heat stress may result in biophysical changes at CLAMP-associated loop anchors that regulate the liquid-liquid phase separation properties of bound proteins. We have recently shown that CLAMP itself can undergo a phase transition (Ray et al. 2024). Several reports suggest that increases in temperature influence phase separation properties of disordered proteins with prion-like domains (PrLD) (Rawat et al. 2021; Zhang et al. 2022; Hutin et al. 2023; Dignon et al. 2019; Chu et al. 2021; Chen et al. 2022). Therefore, it is possible that HS induces changes in the phase separation properties of CLAMP at the loop anchors. Also, CLAMP interacts with the Negative Elongation Factor (NELF) that forms condensates under stress to drive transcriptional repression and regulates Pol II pausing (Rawat et al. 2021; Urban et al. 2017). In the future, it will be interesting to define how the CLAMP-NELF interaction and the relationship between CLAMP and Ibf1/2 regulates HS-induced transcriptional repression. The presence of a different combination of chromatin-bound factors at different CLAMP-bound target genes may regulate the phase transition properties of the locus, which could determine whether a gene will be activated or repressed upon HS. However, further experiments need to be done to test this hypothesis using *in vitro* chromatin models.

Overall, we show that the GA-binding TF CLAMP is a key DNA-binding factor that promotes

transcriptional repression after HS more strongly than activation. Insight into transcriptional repression is important because the mechanism of HS-induced transcriptional repression has remained much less understood than HS-induced activation across species. We also generate new high-resolution 3D and gene expression data sets that can be used for developing combinatorial predictive models of how gene expression is rapidly modulated under different conditions. However, future work is necessary to define the combinatorial context-specific interactions between all of the factors we identified and other unknown factors that determine whether a gene will become activated or repressed upon heat stress.

3.5 Materials and methods

3.5.1 Fly strains and husbandry

Drosophila melanogaster fly stocks were maintained at 24°C on standard corn flour sucrose media. Fly strains used: *fkf-GAL4* (Bloomington, #78061), *UAS-CLAMP RNAi*[val20] (Bloomington, #57163). These were crossed to obtain the desired genotype, *fkf-GAL4>UAS-CLAMP RNAi*.

3.5.2 Cell culture

Kc cells were maintained at 25°C in Schneider's media supplemented with 10% Fetal Bovine Serum and 1.4X Antibiotic-Antimycotic (ThermoFisher Scientific, USA). Cells were passaged every 3 days to maintain an appropriate cell density.

3.5.3 Polytene chromosome squashes and immunostaining after HS

Third instar larvae from undriven *UAS-CLAMP RNAi* and *fkf-GAL4>UAS-CLAMP RNAi* were incubated for 40 min at 37°C in a water bath inside a microfuge tube plugged with a soft plug (Lakhotia et al. 2012) and dissected immediately to pull out the salivary gland. Polytene chromosome squashes were prepared using dissected salivary glands as previously described in Reider et al. (2017). We stained polytene chromosomes with rabbit anti-CLAMP (1:1000, SDIX) and mouse anti-RNA pol II (1:500, ab817, Abcam) antibodies. For detection, we used all Alexa Fluor secondary antibodies against rabbit and mouse at a concentration of 1:200 and visualized slides at 40X on a Zeiss Axioimager M1 Epifluorescence upright microscope with the AxioVision version 4.8.2 software.

3.5.4 Western Blotting to validate *clamp* dsRNA treatment for *clamp* RNAi

Total protein from the samples (1ml of the cell culture): 3 replicates of *clamp* RNAi and *gfp* RNAi (con) after HS (HS) and without HS (NHS) treated Kc cells for SLAM-seq was extracted by homogenizing cell pellet in the lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% SDS, 0.5× protease inhibitor) using a small pestle. The NHS Kc cell samples treated with *clamp* dsRNA and *gfp* dsRNA for *clamp* RNAi and *gfp* RNAi used for Micro-C experiment were also used to extract total protein to test Clamp level in *clamp* RNAi vs *gfp* RNAi Kc cell samples. After a 5 min incubation at room temperature, cleared the samples by centrifuging at room temperature for 10 min at 14,000×g. To blot for CLAMP and Actin, 5 µg of total protein was run on a Novex 10% Tris-Glycine precast gel (Life Technologies). Protein was transferred to PVDF membranes using the iBlot transfer system (Thermo Fisher Scientific) and probed the membranes for CLAMP (1:1000, SDIX), Tubulin (1:5000, Abcam), and Actin (1:5000) antibodies using the Western Breeze kit following the manufacturer's protocol (Thermo Fisher Scientific). Proteins detected on X-ray film (Fig. S3.1) using chemifluorescence detection method as mentioned in the kit protocol using AP-tagged secondary antibody and CDP-star substrate provided with the kit (Western Breeze kit, Thermo Fisher scientific, USA)

3.5.5 SLAM-seq

15 µg each of *clamp* dsRNA and *gfp* dsRNA were used for *clamp* RNAi and *gfp* RNAi (con), respectively, per T25 flask. Kc cells were incubated with dsRNA in FBS minus media for 45 minutes and allowed to grow in media supplemented with 10% FBS for 6 days before harvesting. dsRNA targeting *gfp* (control) and *clamp* for RNAi have been previously validated and described (Larschan et al. 2012). PCR products were used as a template to generate dsRNA using the T7 Megascript kit (Ambion, Inc., USA), followed by purification with the Qiagen RNeasy kit (Qiagen, USA).

Two replicates of each for *clamp* RNAi and *gfp* RNAi, HS (Heat stress) samples, and four replicates each for *clamp* RNAi and *gfp* RNAi, NHS (NHS) samples were used for SLAM-seq. Just before harvesting, cells were either incubated in 4SU (4-thiouridine) provided with SLAMseq Anabolic Kinetics Module (Cat. No. 061.24, Lexogen) at RT or at 37°C for one hour in the dark as per the manufacturer's instructions. Cells were immediately harvested and the pellet was resuspended in 1 mL of Trizol (Invitrogen, USA). RNA isolated in the dark and S4U-labeled transcripts are alkyl-

lated with Iodoacetamide as per manufacturer instruction (SLAMseq Explorer and Kinetics Kit user guide, Lexogen). Resulting modified total RNA was used for downstream NGS library preparation using QuantSeq 3'mRNA-Seq Library Prep Kit for Illumina (FWD) (Cat. No. 015, Lexogen). Libraries sequenced in Illumina HiSeq 4000 with 150x2 bp pair-end sequencing. For analysis Read 2 for the pair-end sequencing was discarded while performing downstream analysis since quality of Read 2 is very low due to the poly(T) stretch at the beginning of Read 2. Data submitted to Gene Expression Omnibus (GSE253645).

3.5.6 Micro-C

Two replicates each for *clamp* RNAi and *gfp* RNAi, HS (Heat stress) samples and for *clamp* RNAi and *gfp* RNAi, NHS (NHS) samples were used for Micro-C. Kc cells (2×10^6) were incubated with 15 μ g each of *clamp* dsRNA and *gfp* dsRNA (same as used for SLAM-seq experiment), used for *clamp* RNAi and *gfp* RNAi (con), respectively, per well of a 6-well plate in FBS minus media for 45 minutes and allowed to grow in media supplemented with 10% FBS for 6 days before harvesting. Before harvesting, HS (Heat stress) samples were incubated at 37°C for one hour. 3×10^6 cells were used per Micro-C reaction. Cells were aliquoted into 1.5 ml tubes, spun at 500g in a swinging bucket rotor for 5 minutes to pellet the cells, which were placed in -80°C after removing the supernatant for 1 hour as per the manufacturer's protocol (Dovetail Micro-C Kit user guide version 2.0). Next, cells were thawed at RT, resuspended in 200 μ L of 1X PBS (Thermo Fisher Scientific, 10010023), and crosslinked using 0.3M DSG and 37% formaldehyde as per the manufacturer's protocol (Dovetail Micro-C Kit user guide version 2.0). Dovetail Micro-C Kit (Catalog # 21006), the Dovetail Library Module for Illumina (Catalog # 25004), and the primer set for NGS library preparation (Dovetail Dual Index Primer Set #1 for Illumina, Catalog # 25010) were used for digestion, proximity ligation, library preparation, ligation capture, and amplification. Libraries were subjected to Illumina 150x2 bp pair-end sequencing with 80X coverage. Data submitted to Gene Expression Omnibus (GSE297379).

3.5.7 Hi-ChIP

Kc Cells were allowed to grow to confluence and harvested. Before harvesting, HS (Heat stress) samples were incubated at 37°C for one hour. 10×10^6 cells were used per HiChIP reaction, and two replicates each for NHS (non-heat stress) and HS (Heat stress). Crosslinking was per-

formed using 0.3M DSG and 37% formaldehyde as per the manufacturer's protocol (Dovetail HiChIP MNase Kit user guide). Dovetail HiChIP MNase Kit (Catalog # 21007), which includes the Dovetail Library Prep Kit and the primer set for NGS library preparation, was used for digestion, lysate preparation, chromatin immunoprecipitation, proximity ligation, and library preparation. Rabbit anti-CLAMP (5µg/sample, SDIX) was used for the immunoprecipitation step. Libraries were subjected to Illumina 150x2 bp pair-end sequencing. Data submitted to Gene Expression Omnibus (GSE253644).

3.5.8 Computational analysis

SLAM-seq Analysis

The nf-core/slamseq (v1.0.0) pipeline was used to process and analyze the SLAM-seq data (Neumann et al. 2019; Ewels et al. 2020). The following parameters were used in the pipeline: -profile singularity, -genome BDGP6, -read-length 150, -trim5 12, -polyA 4, -multimappers true, -quantseq false, -endtoend false, -min-coverage 2, -var-fraction 0.2, -conversions 1, -base-quality 27, -pvalue 0.05, -skip-trimming false, -skip-deseq2 false. Adapter trimming was performed using Trim Galore (v0.6.5). Conversion-aware mapping, alignment filtering, multimapper recovery, SNP calling, read quantification, gene-level quantification collapsing, QC stats, and result summarizations were all performed using SlamDunk (v0.4.3)(Neumann et al. 2019). Differentially expressed genes were then determined using DESeq2 (v1.22.1) (Love et al. 2014).

Micro-C analysis

Micro-C raw data was analyzed using the HiC-Pro toolkit (v3.1.0) (Servant et al. 2015) with default parameters. Normalized contact maps deriving from HiC-Pro were converted to the cooler format using utility scripts (<https://github.com/nservant/HiC-Pro/tree/master/bin/utils>) in the HiC-Pro toolkit (v3.1.0) (Servant et al. 2015). Subsequently, the cooler contact maps for each condition were used to identify differential chromatin interactions using the HiCcompare library (v1.24.0) (Stansfield et al. 2018) in Bioconductor (3.18). Significant differential chromatin interactions ($P < 0.01$) were used in a range of analyses throughout the study. Differential loops within our statistical cutoff were visualized in a MA plot in Python (3.8) using matplotlib (v3.7.5) and seaborn (v0.13.2). Furthermore,

the bedtools suite (v2.31.0) was used to perform overlaps (bedtools pairtobed) with dm6 genes, determining the number of genes that were localized within loops and their anchors, respectively. The Mann-Whitney test was used to determine the statistical significance of differential loop sizes between our studied conditions. Moreover, differential loops were visualized and compared with other datasets in the study using the coolbox API (v0.3.9) (Xu et al. 2021).

Hi-ChIP Analysis

Hi-ChIP raw data was analyzed using the Dovetail genomics Hi-ChIP documentation (<https://hichip.readthedocs.io/en/latest/index.html>). Sequences were mapped to the dm6 genome using the Burrows-Wheeler Aligner (v0.7.17) (Li et al. 2009) passing the following arguments: mem, -5, -S, -P, -T0; default values were used for all other settings. Using the parse module of pairtools (v0.3.0), valid ligation events were recorded; the following parameters were changed from default values: `--min-mapq 40`, `--walks-policy 5unique`, `--max-inter-align-gap 30`. PCR duplicates were removed using the dedup module of pairtools (v0.3.0). Subsequently, the split module of pairtools was used to generate the final bam and pair files. The samtools (v1.17) sort module was used to sort the bam files. To determine the quality of the proximity ligation events, scripts generated by the Dovetail Genomics team were utilized, which are deposited in the following GitHub repository (<https://github.com/dovetail-genomics/HiChIP>).

The cooler suite (v0.9.3) (Abdennur and Mirny 2020) was used to generate contact maps from pair files, and the coolbox API (v0.3.9) (Xu et al. 2021) was used to visualize the contact maps and bigwig tracks. To call loops, the flexible FitHiChIP tool (v10.0) (Bhattacharyya et al. 2019) was used as resolutions are easily adjustable, and it can detect different categories of 3D chromatin interactions. Furthermore, since FitHiChIP required 1D peaks from the same tissue and conditions, MACS2 (v2.2.9.1) was used to call peaks on the Hi-ChIP data. Differential loops between conditions were determined using edgeR in custom scripts located in the FitHiChIP repository (<https://github.com/ay-lab/FitHiChIP>) and were visualized in a MA plot in Python (v3.8) using matplotlib (v3.7.5) and seaborn (v0.13.2). Differential loops were visualized using APA plots with the coolpup.py (Hsieh et al. 2020) suite (v1.0.0), displaying average signal over loop focal points. The central pixel within the APA plot is the central bin (2.5 kb), which contains the differential chromatin contact. The total size of the region displayed, including the central bin, is represented within the

figures legends or within plot. Scores represented in the top left corners of the APA plot correspond to the central pixel intensity relative to the background. Furthermore, the bedtools suite (v2.31.0) was used to perform overlaps (bedtools pairtobed) with dm6 genes, determining the number of genes that were localized within loops and their anchors, respectively. The Mann-Whitney test was used to determine the statistical significance of differential loop sizes between our studied conditions.

Integration of Micro-C, Hi-ChIP, and SLAM-seq datasets

To overlap Micro-C and HiChIP (3D genomics) with SLAM-seq (nascent RNA sequencing), we first had to transform the output of each technique into a mutually comparable format. Therefore, to analyze all three datasets, Micro-C and HiChIP differential loops were annotated with FlyBase genes, and SLAM-seq differential transcripts were converted to FlyBase genes. Micro-C and HiChIP differential loop anchors were annotated using a custom Python script and the bedtools suite (v2.31.0). In addition, the entire loop span was annotated using the bedtools suite (v2.31.0), collecting all FlyBase genes that were localized at the loop anchors and within the loop. Using this method, we derived gene lists from the differential loop analyses of Micro-C and HiChIP, which we could compare with the SLAM-seq differential gene lists. All overlaps presented were visualized using the matplotlib-venn (v0.11.9) Python package, and Fisher's Exact test combined with FDR correction was used to determine statistical significance between all gene lists (Table S3.1). The coolbox API (v0.3.9) (Xu et al. 2021) was used to visually compare all datasets in the study, observing differential contact maps, chromatin loops, ChIP signal, nascent RNA signal, and genes within a specified region.

ChIP-seq Analysis

Raw data (GSE118047, GSE30740, GSE36374, GSE36393, GSE54529, GSE63518, GSE80702, GSE89244) (Li et al. 2015; Gutierrez-Perez et al. 2019; Wood et al. 2011; Yang et al. 2013; Kellner et al. 2012; Van Bortle et al. 2012; Van Bortle et al. 2014; Cubenas-Potts et al. 2017; Rowley et al. 2017) was filtered using Trim Galore (v0.5.0), then aligned to the dm6 genome using the mem module of the Burrows-Wheeler Aligner (v0.7.17) with default settings. BigWig tracks were produced using the bamCoverage module in the deeptools suite (v3.2.1) with the following pa-

rameters: `--binSize 50, --normalizeUsing CPM, --extendReads 200, --exactScaling, --centerReads, --blackListFileName dm6-blacklist.bed`. The respective BigWig tracks were visualized using the coolbox API (v0.3.9) and the deeptools suite (v3.2.1) over regions of interest.

3.6 Competing Interest Statement

The authors declare no conflict of interest.

3.7 Funding

This work was supported by R35GM126994 to E.N.L. from NIH (National Institutes of Health (NIH)). J.L.A. is funded by the HHMI Gilliam Scholar Program. S.M.L. was funded by Brown University's Undergraduate Teaching and Research Awards (UTRAs). K.C. is a Blavatnik scholar.

3.8 Author Contributions

J.L.A., M.R., E.N.L. planned experiments, analyzed results, and wrote the manuscript. J.L.A., M.R., and K.C. performed micro-C experiments. M.R. and J.L.A. performed the HiChIP experiments. M.R. and J.D. performed the SLAM-seq experiments. J.D. performed computational analysis of the SLAM-seq data. J.L.A. and A.A. performed computational analysis of micro-C and HiChIP data, generated and integrated both data sets with available ChIP data sets, and generated SLAM-seq data sets. J.D. reviewed and edited the manuscript. S.M.L. performed the polytene chromosome squashes and immunostaining after HS.

3.9 Supplemental Figures

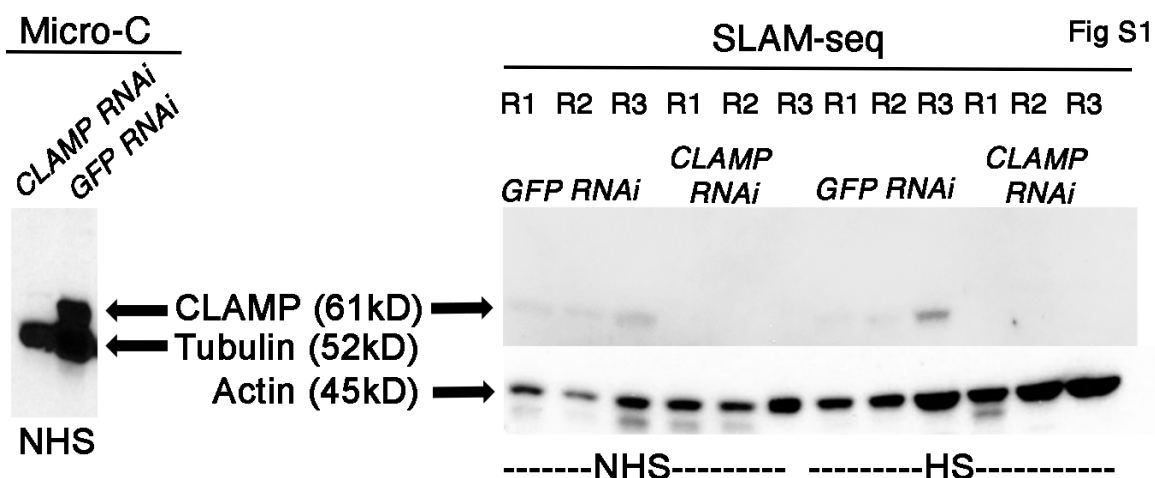


Figure S3.1: Photomicrograph showing validation of CLAMP dsRNA treatment, lowering CLAMP protein level in Kc cells.

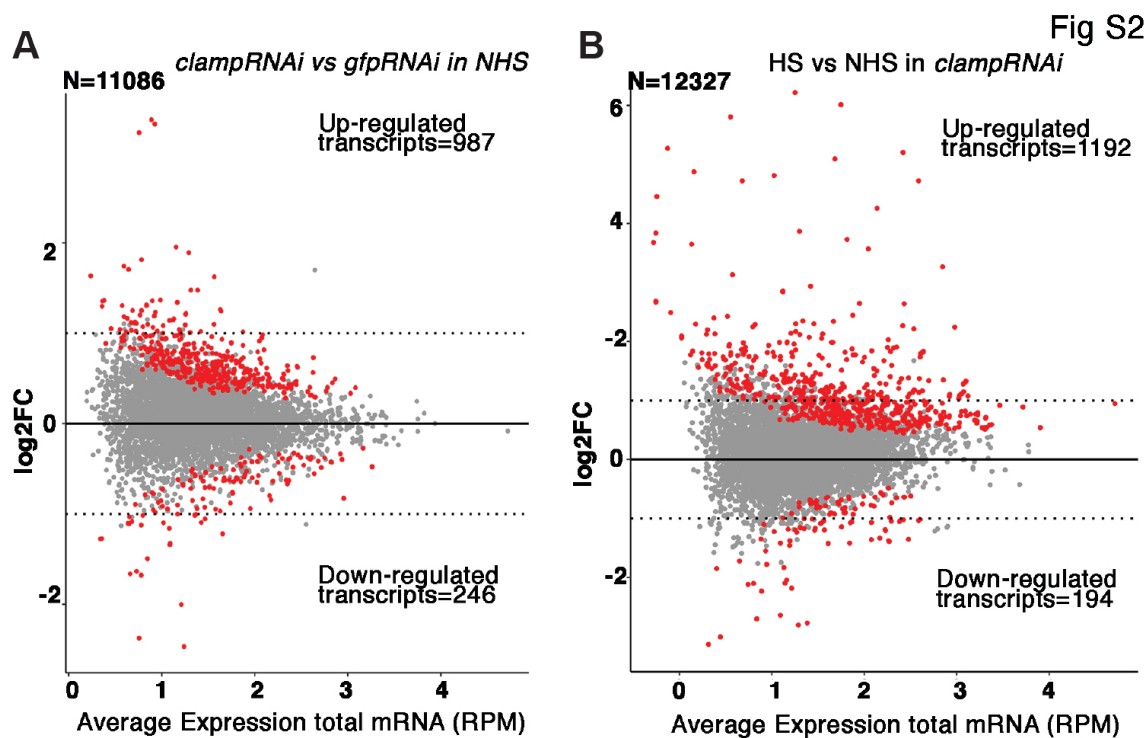


Figure S3.2: **CLAMP functions more frequently as a repressor than an activator in both NHS and HS conditions.** (A-B) MA plots show more up-regulated nascent transcripts (positive red dots, $p < 0.05$, $\log_2FC \geq 2$) in both *clamp* RNAi-treated cells under NHS (no heat stress) and HS (heat stress) conditions.

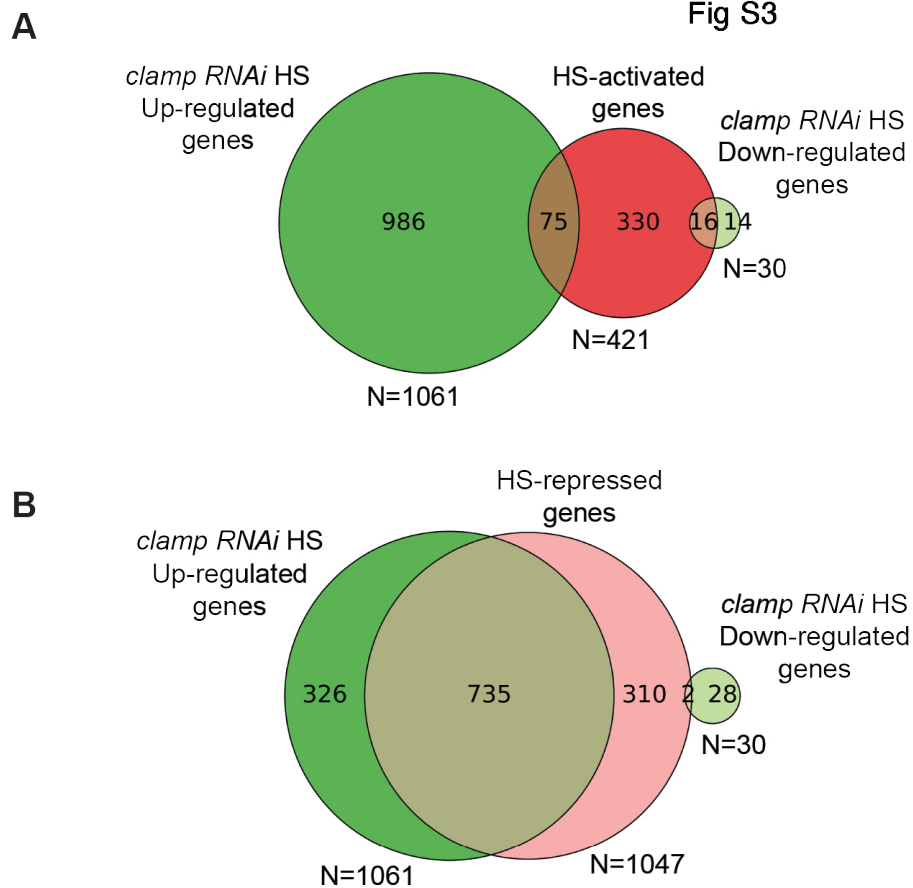


Figure S3.3: **CLAMP-dependent HS-repressed genes significantly overlap with HS-repressed genes.** (A-B) Venn diagrams comparing HS-activated, N=421 (A), and HS-repressed, N=1047 (B) genes with *clamp* RNAi-dependent HS-upregulated (N=1061) and downregulated (N=30) genes. *clamp* RNAi upregulated genes significantly overlap with HS-repressed genes ($p=0$, Fisher's test with FDR correction).

Fig S4

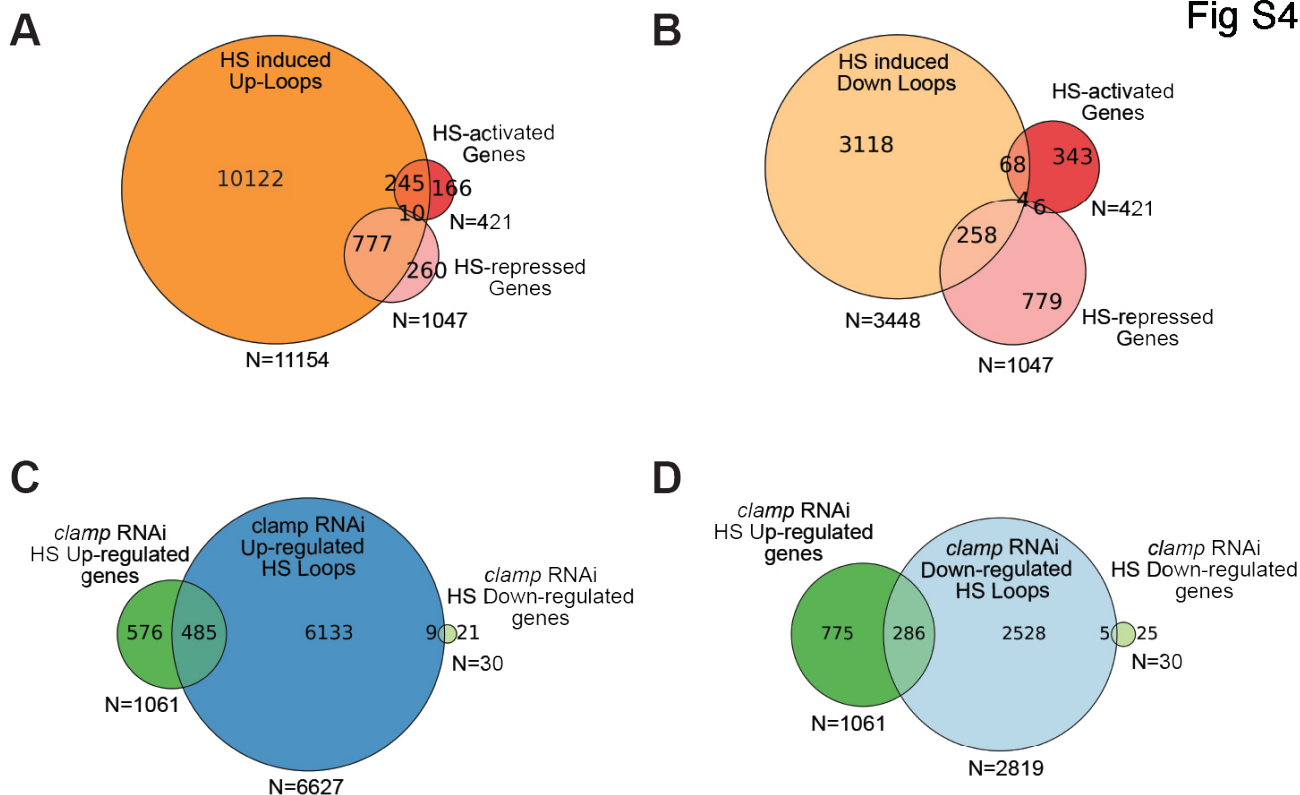


Figure S3.4: **Many CLAMP-dependent repressed genes are present at the loop anchors of CLAMP-dependent HS-associated chromatin loops lost after HS.** (A-D) Venn diagrams comparing HS-associated chromatin loop changes (genes at the loop anchors) and HS-activated and repressed genes (A and B); as well as CLAMP-dependent chromatin loop changes during HS with CLAMP-dependent HS-induced transcriptional changes (C and D), showing significant (Fisher's Test) overlap between HS-associated Up loops (A) and Down loops (B) with HS-repressed genes; and CLAMP-dependent HS-associated Down loops (C, *clamp* RNAi up-regulated HS-loops) and Up loops (D, *clamp* RNAi down-regulated HS-loops) with CLAMP-dependent HS-repressed genes (*clamp* RNAi HS up-regulated genes).

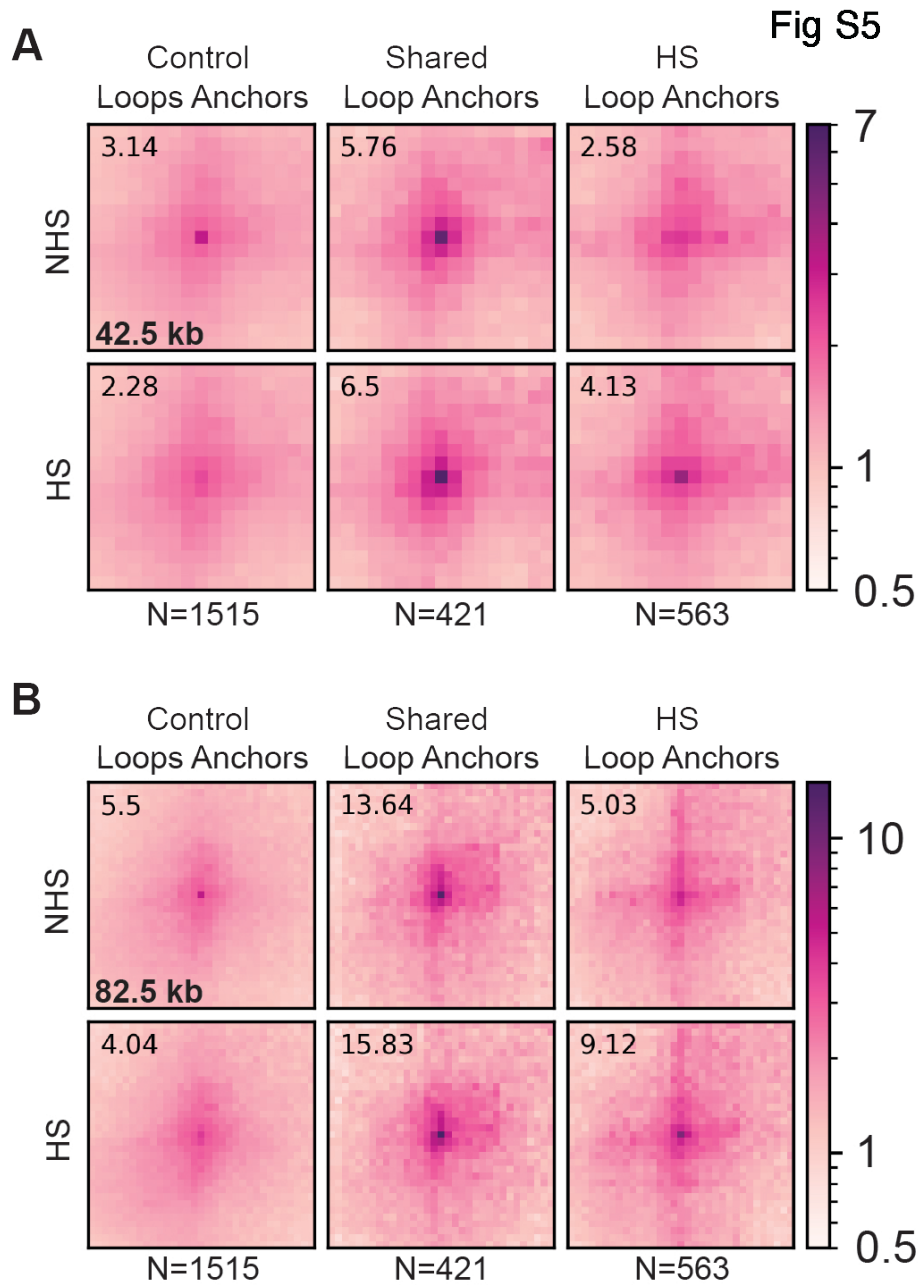


Figure S3.5: **CLAMP-HiChIP identifies CLAMP-bound chromatin loops in control and HS cells.** (A-B) APA plots of CLAMP HiChIP loops centered on three classes of significant loop interactions identified by CLAMP HiChIP in both NHS and HS conditions: 1) Loops identified only in NHS conditions (control loops, N=1515); 2) Loops identified in both NHS and HS conditions (shared loops, N=421) and 3) Loops identified only in HS conditions (HS loops, N=563). The value listed indicates the enrichment of the center pixel compared to the background, at ~40 kb (A) and ~80 kb (B) resolution.

Fig S6

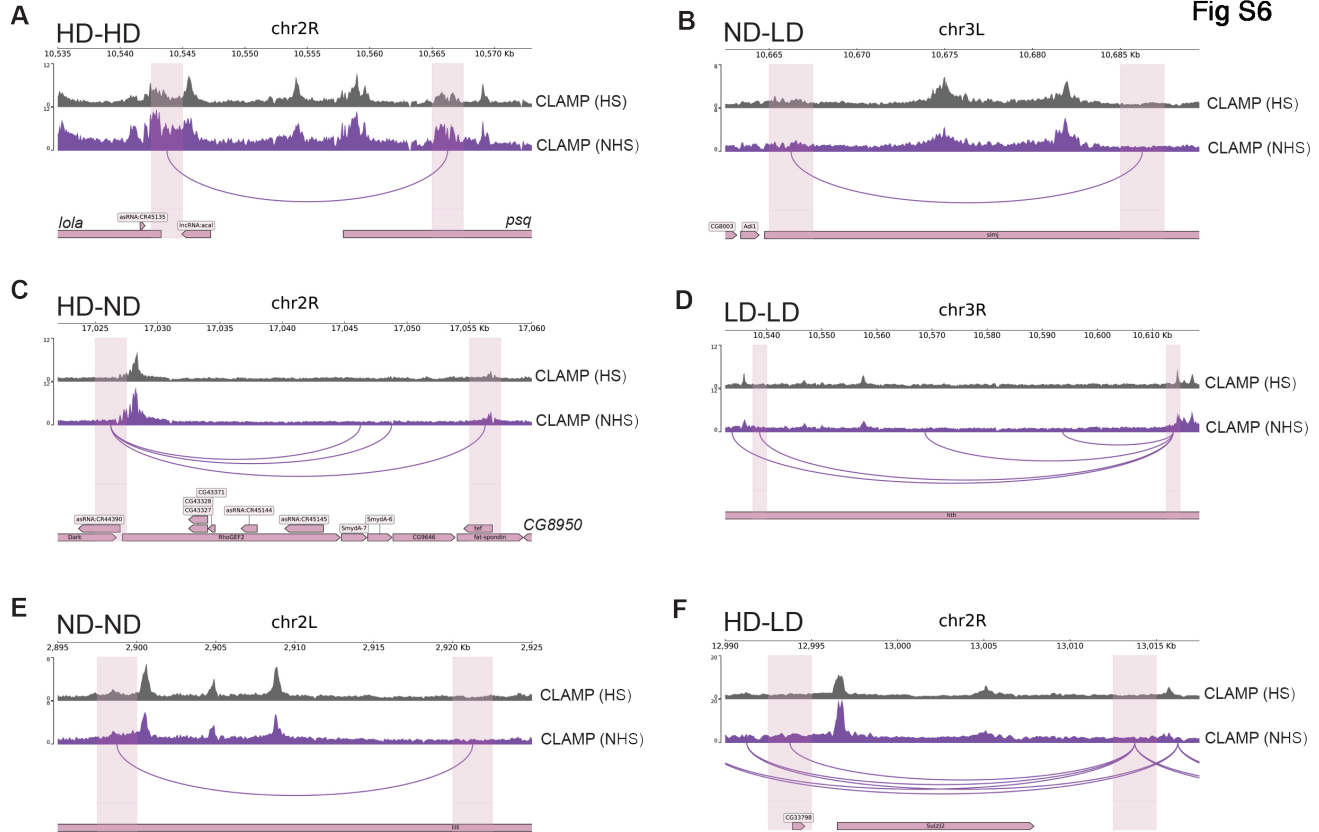


Figure S3.6: Most CLAMP-associated changes upon heat stress are three-dimensional and not associated with differences in CLAMP occupancy. (A-F) Example tracks for each class of CLAMP-associated loops lost and gained after HS. The grey tracks show the occupancy of CLAMP upon HS, whereas the purple tracks show CLAMP occupancy in normal conditions (NHS). A purple line connects loop anchor endpoints. Highlighted regions correspond to 2.5 kb anchors. The classes of CLAMP-associated loops shown are: **(A)** HD-HD, in which there is a change in both three-dimensional contacts and CLAMP occupancy at both ends of the loop anchor upon HS; **(B)** ND-LD in which there is a change in 3D contacts and only a slight change in CLAMP occupancy at one of the ends of the loop anchor upon HS; **(C)** HD-ND in which there is a change in 3D contacts and occupancy of CLAMP at one end of the loop anchor upon HS; **(D)** LD-LD in which there is a change in 3D contacts and only a slight change in CLAMP occupancy at both of the ends of the loop anchor upon HS. **(E)** ND-ND in which there is only a change in 3D contacts and no change in CLAMP occupancy at either end of the loop anchor upon HS; **(F)** HD-LD in which there is a change in 3D contact at one end of the loop anchor and only slight change in CLAMP occupancy at another end of the loop anchor upon HS.

3.10 Supplementary Tables

Table S3.1: List of statistical tests and their results listed for the Venn diagrams generated for analysis of micro-C, HiChIP, and SLAM-seq data sets.

Group 1	Group 2	Over- lap	Group 1 size	Group 2 size	Raw p-value	Adj. p-value	Signifi- cant
Figure 1D Fisher's test + FDR correction							
HS-activated transcripts	<i>clamp RNAi HS</i> <i>Up-regulated</i> <i>transcripts</i>	190	995	4028	1.750E-13	2.625E-13	Yes
HS-activated transcripts	<i>clamp RNAi HS</i> <i>Down-regulated</i> <i>transcripts</i>	29	995	58	7.447E-30	2.234E-29	Yes
<i>clamp RNAi HS</i> <i>Up-regulated</i> <i>transcripts</i>	<i>clamp RNAi HS</i> <i>Down-regulated</i> <i>transcripts</i>	0	4028	58	1.000E+00	1.000E+00	No
Figure 1E Fisher's test + FDR correction							
HS-repressed transcripts	<i>clamp RNAi HS</i> <i>Up-regulated</i> <i>transcripts</i>	2660	3547	4028	0	0	Yes

Continued on next page...

Group 1	Group 2	Over- lap	Group 1 size	Group 2 size	Raw p-value	Adj. p-value	Signifi- cant
HS-repressed transcripts	<i>clamp RNAi HS</i> <i>Down-regulated</i> <i>transcripts</i>	2	3547	58	0.982895	1	No
<i>clamp RNAi HS</i> <i>Up-regulated</i> <i>transcripts</i>	<i>clamp RNAi HS</i> <i>Down-regulated</i> <i>transcripts</i>	0	4028	58	1	1	No
Figure 2D Mann–Whitney test + FDR correction							
<i>clamp RNAi HS</i> <i>Down-reg. Loops</i>	<i>clamp RNAi HS</i> <i>Up-reg. Loops</i>	—	—	—	3.966E-70	3.97E-70	Yes
<i>clamp RNAi HS</i> <i>Down-reg. Loops</i>	HS induced Down Loops	—	—	—	0.000E+00	0.000E+00	Yes
<i>clamp RNAi HS</i> <i>Down-reg. Loops</i>	HS induced Up Loops	—	—	—	0.000E+00	0.000E+00	Yes
<i>clamp RNAi HS</i> <i>Up-reg. Loops</i>	HS induced Down Loops	—	—	—	0.000E+00	0.000E+00	Yes
<i>clamp RNAi HS</i> <i>Up-reg. Loops</i>	HS induced Up Loops	—	—	—	0.000E+00	0.000E+00	Yes

Continued on next page...

Group 1	Group 2	Over- lap	Group 1 size	Group 2 size	Raw p-value	Adj. p-value	Signifi- cant
HS induced Down Loops	HS induced Up Loops	—	—	—	—	5.63E-186	Yes
Figure 2F Fisher's test + FDR correction							
HS induced Up Loops	HS-activated genes	344	14132	421	8.906E-02	1.336E-01	No
HS induced Up Loops	HS-repressed genes	899	14132	1047	2.496E-09	7.487E-09	Yes
HS-activated genes	HS-repressed genes	10	421	1047	9.998E-01	9.998E-01	No
Figure 2G Fisher's test + FDR correction							
HS induced Down Loops	HS-activated genes	323	13223	421	0.098601	0.147901	No
HS induced Down Loops	HS-repressed genes	825	13223	1047	0.000084	0.000253	Yes
HS-activated genes	HS-repressed genes	10	421	1047	0.999819	0.999819	No
Figure 2H Fisher's test + FDR correction							
<i>clamp RNAi Up-reg. HS Loops</i>	<i>clamp RNAi HS Up-reg. Genes</i>	717	10889	1061	0.000002	0.000005	Yes
<i>clamp RNAi Up-reg. HS Loops</i>	<i>clamp RNAi HS Down-reg. Genes</i>	21	10889	30	0.201558	0.302336	No

Continued on next page...

Group 1	Group 2	Over- lap	Group 1 size	Group 2 size	Raw p-value	Adj. p-value	Signifi- cant
<i>clamp RNAi HS</i> <i>Up-reg. Genes</i>	<i>clamp RNAi HS</i> <i>Down-reg. Genes</i>	0	1061	30	1	1	No
Figure 2I Fisher's test + FDR correction							
<i>clamp RNAi</i> <i>Down-reg. HS Loops</i>	<i>clamp RNAi HS</i> <i>Up-reg. Genes</i>	445	5506	1061	1.430E-15	4.290E-15	Yes
<i>clamp RNAi</i> <i>Down-reg. HS Loops</i>	<i>clamp RNAi HS</i> <i>Down-reg. Genes</i>	14	5506	30	4.885E-02	7.327E-02	No
<i>clamp RNAi HS</i> <i>Up-reg. Genes</i>	<i>clamp RNAi HS</i> <i>Down-reg. Genes</i>	0	1061	30	1.000E+00	1.000E+00	No
Figure 3D Mann-Whitney test + FDR correction							
CLAMP assoc. HS induced Up Loops	CLAMP assoc. HS induced Down Loops	—	—	—	0.035489	0.035489	Yes
Figure 3E Fisher's test + FDR correction							
HS induced Down Loops	CLAMP assoc. HS induced Down Loops	1459	13223	1532	4.208E- 116	1.262E- 115	Yes
HS induced Down Loops	HS-repressed genes	825	13223	1047	8.424E-05	8.424E-05	Yes

Continued on next page...

Group 1	Group 2	Over- lap	Group 1 size	Group 2 size	Raw p-value	Adj. p-value	Signifi- cant
CLAMP assoc. HS induced Down Loops	HS-repressed genes	173	1532	1047	6.397E-18	9.595E-18	Yes
Figure 3F Fisher's test + FDR correction							
HS induced Down Loops	CLAMP assoc. HS induced Down Loops	1459	13223	1532	4.208E- 116	1.262E- 115	Yes
HS induced Down Loops	<i>clamp RNAi HS</i> <i>Up-reg. Genes</i>	834	13223	1061	1.381E-04	1.381E-04	Yes
CLAMP assoc. HS induced Down Loops	<i>clamp RNAi HS</i> <i>Up-reg. Genes</i>	179	1532	1061	1.712E-19	2.567E-19	Yes
Figure 4A Fisher's test + FDR correction							
<i>clamp RNAi Up-reg.</i> <i>HS Loops</i>	CLAMP assoc. HS induced Down Loops	1142	10889	1532	2.93E-32	8.78E-32	Yes
<i>clamp RNAi Up-reg.</i> <i>HS Loops</i>	HS-repressed + <i>clamp RNAi HS</i> Up-reg. Genes	494	10889	735	1.55E-04	1.55E-04	Yes
CLAMP assoc. HS induced Down Loops	HS-repressed + <i>clamp RNAi HS</i> Up-reg. Genes	130	1532	735	5.65E-16	8.47E-16	Yes

Continued on next page...

Group 1	Group 2	Over- lap	Group 1 size	Group 2 size	Raw p-value	Adj. p-value	Signifi- cant
Figure S2A Fisher's test + FDR correction							
HS-activated genes	<i>clamp RNAi HS</i> <i>Up-reg. genes</i>	75	421	1061	9.15E-17	1.37E-16	Yes
HS-activated genes	<i>clamp RNAi HS</i> <i>Down-reg. genes</i>	16	421	30	1.59E-18	4.79E-18	Yes
<i>clamp RNAi HS</i> <i>Up-reg. genes</i>	<i>clamp RNAi HS</i> <i>Down-reg. genes</i>	0	1061	30	1	1	No
Figure S2B Fisher's test + FDR correction							
HS-repressed genes	<i>clamp RNAi HS</i> <i>Up-reg. genes</i>	735	1047	1061	0.00E+00	0.00E+00	Yes
HS-repressed genes	<i>clamp RNAi HS</i> <i>Down-reg. genes</i>	2	1047	30	5.50E-01	8.30E-01	No
<i>clamp RNAi HS</i> <i>Up-reg. genes</i>	<i>clamp RNAi HS</i> <i>Down-reg. genes</i>	0	1061	30	1	1	No

– Not applicable (Mann–Whitney; no overlap metric). Adj. = FDR-adjusted. *Italic* = RNAi experimental condition.

Chapter 4: Leveraging graph neural networks to classify GA-repeat motifs using epigenomic features and 3D chromatin structure

Joseph Aguilera¹, Pranav Mahableshwarkar¹, Sarah Gunasekera¹, Romer Miranda¹, Tuan Pham¹, Cynthia Zhang¹, Corinn Tiffany¹, and Erica Larschan¹

I wrote this chapter with editorial feedback from Dr. Erica Larschan.

This study is still in progress. I aim to submit this Chapter for publication at *PLOS Computational Biology* by fall of 2026.

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

4.1 Abstract

The GA-binding pioneer transcription factors CLAMP and GAF bind to a subset of GA-repeat motifs distributed throughout the *Drosophila* genome, yet the features that distinguish bound from unbound GA-repeats remain poorly understood. Here, we leveraged ubiquitous, genome-wide histone modification signal and high-resolution three-dimensional (3D) chromatin interaction data derived from male *Drosophila* S2 cells to characterize the chromatin environment present at every GA-repeat motif in the genome. We hypothesized that the inclusion of 3D chromatin interaction data would be critical in developing robust models capable of distinguishing between classes of GA-repeat motifs. To test this hypothesis, we employed three machine and deep learning algorithms: logistic regression (LR), a multi-layer perceptron (MLP), and a graph convolutional neural network (GCN). The GCN model, which integrates both epigenomic features and 3D chromatin structure, substantially outperformed the LR and MLP models, achieving an AUROC of 0.98 compared to 0.83 and an AUPR of 0.70 compared to 0.51. Using GNNExplainer, we derived feature importance scores revealing that the H3K4me1, H3K4me2, H3K27me3, and H4K16ac histone modifications are critical in distinguishing between GA-repeats throughout the genome. Our findings demonstrate that the chromatin environment surrounding GA-repeat motifs, particularly when captured in the context of 3D chromatin organization, possesses unique epigenomic signatures which generate new hypotheses about factors that drive pioneer transcription factor occupancy.

4.2 Introduction

Precise and coordinated regulation of gene expression is essential for the viability of all organisms (Lee and Young 2013). Chromatin domains coordinate gene expression by concentrating key factors at specific genomic locations in a context-specific way to concordantly activate or repress groups of genes. Within the nucleus of eukaryotic cells, chromatin is tightly compacted through multiple orders of organization and shares finite nuclear space with many classes of transcriptional regulators, including DNA-binding transcription factors (TFs) and large transcription complexes (Bonev and Cavalli 2016; Zaret 2020). TFs recruit large transcription complexes to facilitate spatially precise and temporally dynamic transcriptional repression and activation of specific groups

of genes (Zaret 2020; Drouin 2014). A fundamental question in chromatin biology is how TFs with similar DNA-binding preferences achieve distinct regulatory outcomes at different genomic loci. This question is particularly relevant for pioneer TFs, which are capable of binding compacted chromatin and facilitating subsequent binding of other factors to establish new regulatory states.

In *Drosophila melanogaster*, the GA-binding pioneer TFs CLAMP (Chromatin-Linked Adaptor for MSL Proteins) and GAF (GAGA Factor, encoded by the *Trithorax-like* gene) recognize GA-repeat motifs distributed throughout the genome (Fuda et al. 2015; Kuzu et al. 2016). These factors play central roles in many biological processes including dosage compensation and the heat stress response (Larson et al. 2021; Freund et al. 2024). However, little is known about what factors determine the context-specific regulatory function of each factor when they bind to different genomic locations (Gaskill et al. 2021; Duan et al. 2021). In the context of dosage compensation, CLAMP facilitates the recruitment of the dosage compensation complex (DCC) to the single male X chromosome, mediating the approximately two-fold transcriptional upregulation of X-linked genes (Soruco et al. 2013; Larschan E et al. 2011). Variation within GA-repeat binding sites that accumulated over evolutionary time on the X chromosome promotes the occupancy of CLAMP instead of GAF, and this differential occupancy is critical for X-specific DCC recruitment (Kaye et al. 2018; Kuzu et al. 2016). Despite the importance of these factors, CLAMP and GAF bind to only a subset of the thousands of GA-repeat motifs present in the genome, raising a fundamental question: what distinguishes the GA-repeat motifs at which these transcription factors bind from those at which they do not?

One approach to addressing this question is to characterize the chromatin environment—defined as the combinatorial representation of histone modifications at a specific locus—present at every GA-repeat motif in the genome. Histone modifications are key regulators of transcriptional activity, with specific modifications driving activation (e.g., H3K4me1, H3K4me2, H3K4me3, H4K16ac) or repression (e.g., H3K27me3, H3K9me3) of gene expression (Kouzarides 2007; Bannister and Kouzarides 2011; Gelbart and Kuroda 2009). By extracting genome-wide signal from these key histone modifications in male *Drosophila* S2 cells, where dosage compensation is constitutively active, we can perform a combinatorial analysis of the epigenomic signal present at GA-repeats. However, histone modification signal captured in linear genomic space provides an incomplete picture of the true regulatory environment. Chromatin is organized in three dimensions, and distal

genomic loci can be brought into close spatial proximity through chromatin looping (Bonev and Cavalli 2016). Therefore, the histone modifications that are functionally relevant to a given GA-repeat may be present at distal loci which interact with it in 3D space.

In recent years, advances in machine learning have provided powerful tools for modeling complex biological data. Convolutional neural networks (CNNs), long short-term memory networks (LSTMs), and transformers have been applied to predict transcription factor binding sites with increasing accuracy (LeCun et al. 2015; Hochreiter and Schmidhuber 1997; Vaswani et al. 2017). CNN-based models such as DeepBind and DeepSEA have demonstrated strong performance in predicting binding from DNA sequence alone (Zhou and Troyanskaya 2015; Alipanahi et al. 2015). However, these approaches do not incorporate epigenomic data or three-dimensional chromatin organization, which are essential for understanding the binding activities of transcription factors such as CLAMP and GAF. Graph neural networks (GNNs) offer a natural framework for modeling 3D chromatin interactions as they operate on graph-structured data where nodes represent genomic loci and edges represent chromatin contacts (Kipf and Welling 2016). Notably, GC-MERGE demonstrated that graph convolutional networks can effectively integrate lower resolution Hi-C-derived chromatin interaction maps with histone modification profiles to predict gene expression across human cell lines (Bigness et al. 2022), providing an interpretable framework that quantifies both node-level features and edge-level interactions in the context of 3D genome organization.

Here, we extend this framework to comprehensively characterize the chromatin environment at every GA-repeat motif present throughout the entire *Drosophila* genome. We leveraged ubiquitous, genome-wide signal from key histone modifications that drive activation and repression of genes, derived from male *Drosophila* S2 cells. Critically, we also incorporated high-resolution 3D chromatin interaction data to determine the combination of histone modifications enriched at each GA-repeat in the context of its spatial neighborhood. We hypothesized that the inclusion of 3D chromatin interaction data would be critical in developing robust models aimed at distinguishing between GA-repeat motifs in the genome. To test this hypothesis, we utilized three machine and deep learning algorithms to build a GA-repeat classifier: logistic regression (LR), a multi-layer perceptron (MLP), and a graph convolutional neural network (GCN). The GCN model outperformed both the LR and MLP models, achieving substantial increases in AUROC (0.83 to 0.98) and AUPR (0.51 to 0.70). Using GNNExplainer (Ying et al. 2019), we derived importance scores revealing

unique epigenomic signatures that distinguish between classes of GA-repeats. These signatures capture the enrichment or depletion of specific histone modifications that aggregate in 3D space, providing substantially more information than what non-graph machine learning models could offer. Together, our results demonstrate that the chromatin environment, when captured in three dimensions, provides the discriminative power necessary to distinguish between GA-repeat motifs and reveals the epigenomic features that may underlie context-specific transcription factor occupancy.

4.3 Experimental Design and Methods

4.3.1 Overview of datasets

To characterize the chromatin environment at genome-wide GA-repeat motifs, we leveraged publicly available ChIP-seq data from male *Drosophila* S2 cells for eight histone modifications that represent key active and repressive chromatin states. The activating marks included H3K4me1 (GSM1535991), H3K4me2 (GSM731241), H3K4me3 (GSM1535994), H3K27ac (GSM2259987), H3K36me3 (GSM731243), and H4K16ac (GSM731244), while the repressive marks included H3K9me3 (GSM2487547) and H3K27me3 (GSM2776903). These modifications were selected because they are ubiquitously distributed throughout the genome and collectively capture the major regulatory states underlying transcriptional activation and repression in *Drosophila* (Kharchenko et al. 2011). Notably, H4K16ac is the hallmark modification deposited by the dosage compensation complex on the male X chromosome (Gelbart and Kuroda 2009), making its inclusion particularly relevant to our analysis of GA-repeat classification. ChIP-seq signal for each modification was binned at 1kb resolution across the genome. Transcription factors were excluded from the study as they possess sequence specificity which may bias the models during training.

To incorporate three-dimensional chromatin organization, we utilized the control (*gfp* RNAi) Micro-C data generated in Chapter 2 from male *Drosophila* S2 cells. Micro-C captures chromatin interactions at high resolution (Hsieh et al. 2015), and these data were similarly processed at 1kb resolution to maintain consistency with the histone modification signal. The use of matched cell types across all datasets ensures that both the epigenomic profiles and 3D chromatin interaction maps reflect the same biological context. Together, these datasets comprise the complete set of input features used to construct the feature matrices for all downstream machine learning models.

4.3.2 Generation of target variable

To identify and classify GA-repeat motifs throughout the *Drosophila* genome, we used the HOMER motif discovery suite (Heinz et al. 2010). We first performed *de novo* motif discovery on the high-affinity sites (HAS) of the dosage compensation complex (Alekseyenko et al. 2008), extracting the core 8 bp GA-repeat motif central to the MSL-recognition elements (Kuzu et al. 2016). Using this consensus motif, allowing for a single mismatch to account for CLAMP or GAF binding (see Chapter 2), we then conducted a genome-wide scan to identify all occurrences of the GA-repeat motif across the *Drosophila* genome.

Subsequently, each identified GA-repeat motif was assigned to one of three classes based on its genomic location. Class 1 comprises GA-repeats located on the autosomes, representing motifs that do not regulate X-chromosome dosage compensation. Class 2 comprises GA-repeats located on the X chromosome but outside of the high-affinity sites, representing X-linked motifs that are not direct DCC recruitment sites. Class 3 comprises GA-repeats located within the high-affinity sites, representing the motifs that directly recruit the dosage compensation complex to the X chromosome.

To integrate these classifications with the 1kb resolution feature matrices described above, we binned the entire genome into non-overlapping 1kb windows and assigned each bin the class label of its resident GA-repeat motif. Bins that did not contain a GA-repeat motif were labeled as Class 0, representing genomic regions without GA-repeats. This four-class labeling scheme (Classes 0 to 3) served as the target variable for all downstream classification models. Notably, all models will learn to discriminate only between classes 1 to 3; class 0 was generated to enable the graph structure described below.

4.3.3 Graph construction and data integration

To construct the graph representation of the *Drosophila* genome, we defined each 1kb genomic bin as a node and assigned chromatin interactions as edges connecting nodes in three-dimensional space. Importantly, not all raw chromatin interactions from the Micro-C data were used as edges. Instead, we identified statistically significant chromatin interactions using FitHiC (Ay et al. 2014; Kaul et al. 2020), keeping only those interactions that passed significance thresholds. This filtering

step ensures that the graph structure reflects biologically meaningful 3D contacts rather than background interactions. However, even after filtering for significant interactions, the average number of connections per node was approximately 300, presenting a computational challenge. During graph convolution, aggregating signal across a large neighborhood would dilute the epigenomic features at each node, averaging out biologically meaningful signal. To address this, we used a *Neighbor Sampler* strategy (Hamilton et al. 2017); for each node, the top 20 connections are sampled by their transformed FitHiC p-values. This sampling approach ensures that the most statistically significant chromatin interactions have the highest likelihood of being selected to represent a given node during each convolution. Together, these edge considerations ensure that the GCN operates on a high-confidence graph in which each node's neighborhood captures the most informative 3D chromatin interactions rather than a diluted aggregate of all possible contacts.

4.3.4 Baseline models

To establish baseline performance for the GA-repeat classification task, we implemented two non-graph models: logistic regression (LR) (Pedregosa et al. 2012) and a multi-layer perceptron (MLP) (Paszke et al. 2019). As input, both models receive the histone modification feature vector for each 1kb genomic bin, without any information about 3D chromatin interactions. For both baselines, input features were standardized using a scikit-learn's *StandardScaler* (Pedregosa et al. 2012).

The logistic regression model was implemented using scikit-learn's *LogisticRegression* with a multinomial formulation optimized via the L-BFGS solver. To address class imbalance, we applied the balanced class weighting scheme, which adjusts weights inversely proportional to class frequency. The model was trained for a maximum of 1,000 iterations with a fixed random seed for reproducibility. LR serves as a linear baseline, capturing only the relationships between individual histone modifications without modeling any interactions between features.

The MLP was implemented in *PyTorch* (Paszke et al. 2019) as a fully connected feedforward neural network consisting of three hidden layers, each with 128 units. Each hidden layer was followed by batch normalization, a ReLU activation function, and dropout regularization (0.3) to reduce overfitting. The output layer projects to the number of target classes. The model was trained for 250 epochs using the Adam optimizer with a learning rate of 1×10^{-3} and weight decay of 5×10^{-4} , with a *ReduceLROnPlateau* scheduler that halved the learning rate after 50 epochs

of no improvement. To address class imbalance, we applied inverse frequency class weights to the cross-entropy loss function during training. Model selection was performed based on the best validation AUROC, and training was conducted on GPU with a batch size of 64. Unlike logistic regression, the MLP can model nonlinear relationships between histone modifications through its hidden layers, but like LR, it operates solely on the feature vector and cannot incorporate 3D chromatin structure.

4.3.5 Graph convolutional neural networks

To leverage 3D chromatin organization, we implemented a graph convolutional neural network (GCN) using *PyTorch Geometric* (Fey and Lenssen 2019). In this framework, each 1kb genomic bin is represented as a node with a feature vector corresponding to the histone modification signal, and edges between nodes represent statistically significant chromatin interactions identified by FitHiC, as described above. The GCN architecture consists of three *GCNConv* layers (Kipf and Welling 2016), each with a hidden dimensionality of 128, followed by three fully connected linear layers also with 128 hidden units, and a final output layer projecting to the number of target classes. Dropout regularization ($p = 0.5$) was applied throughout the network, and normalization was enabled within the *GCNConv* layers.

The key distinction of the GCN from the baseline models lies in its message-passing mechanism. At each *GCNConv* layer, each node aggregates the feature representations of its neighboring nodes and updates its own representation accordingly. Over three convolutional layers, each node's representation incorporates information from its three-hop neighborhood in 3D chromatin space, capturing the combinatorial histone modification environment at the loci it contacts in three dimensions.

During training, we employed a *NeighborLoader* with a sampling strategy of 20 neighbors per node per layer, selecting the most significant chromatin interactions based on their transformed FitHiC p-values, as described above. Training was performed using the Adam optimizer with a learning rate of 5×10^{-4} and weight decay of 5×10^{-4} for 50 epochs with a batch size of 256. Inverse frequency class weights were applied to the cross-entropy loss to address class imbalance. The model was trained on GPU, and all experiments were conducted with a fixed random seed for reproducibility.

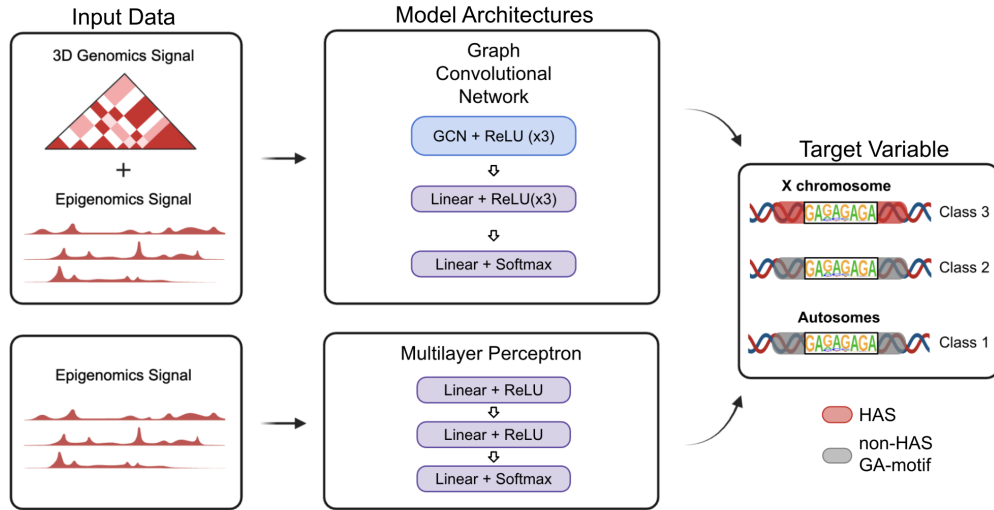


Figure 4.1: **Experimental overview of the GA-repeat classification framework.** The input data consist of two complementary data types derived from male *Drosophila* S2 cells: (top) 3D genomics signal from Micro-C chromatin interaction maps combined with epigenomic signal from eight histone modification ChIP-seq datasets, and (bottom) epigenomic signal alone. The GCN (top) receives both data types as a graph where nodes represent 1kb genomic bins with histone modification feature vectors and edges represent significant chromatin interactions identified by FitHiC. The MLP (bottom) receives only the epigenomic feature vectors without 3D structural information. Both architectures output predictions across three target classes: Class 1 (GA-repeats on autosomes), Class 2 (GA-repeats on the X chromosome outside of high-affinity sites), and Class 3 (GA-repeats within high-affinity sites, HAS).

4.3.6 Evaluation metrics

To evaluate model performance, we used two complementary metrics: the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPR). AUROC measures the ability of a classifier to distinguish between classes across all decision thresholds and is widely used in binary and multi-class classification tasks. However, AUROC can be misleading when applied to highly imbalanced datasets as it is influenced by the true negative rate, which is inflated when the majority class dominates. Since our dataset is highly imbalanced, representing Class 1 (autosomes) substantially more than Classes 2 (X chromosome) and 3 (high-affinity sites), we paired AUROC with AUPR, focusing on precision and recall which is more sensitive to the performance of the minority classes. To further reduce the effects of class imbalance during training, we applied class weights inversely proportional to class frequency. Together, the use of both metrics alongside class weighting provides a more complete and rigorous assessment of model performance across all GA-repeat classes.

4.3.7 Experimental overview

To address the question of what distinguishes GA-repeat motifs throughout the genome, we designed a classification framework that integrates epigenomic and 3D chromatin data (Fig. 4.1). Each 1kb genomic bin was represented as a node in a graph, with eight histone modification ChIP-seq signals as node features and statistically significant Micro-C chromatin interactions as edges. GA-repeat motifs were classified into three categories: Class 1 (autosomal GA-repeats), Class 2 (X-chromosomal GA-repeats outside of high-affinity sites), and Class 3 (GA-repeats within HAS that directly recruit the DCC). We compared three models with increasing complexity: a logistic regression (LR) that captures only linear relationships between histone modifications, a multi-layer perceptron (MLP) that models nonlinear feature interactions but ignores 3D structure, and a graph convolutional neural network (GCN) that aggregates histone modification signal across the 3D chromatin neighborhood of each GA-repeat. This design enables us to directly assess the contribution of 3D chromatin organization to the classification of GA-repeat motifs.

4.4 Results and Discussion

4.4.1 Graph convolutional neural networks outperform baseline models

We first compared the performance of the three classification models on the GA-repeat classification task (Fig. 4.2). The logistic regression (LR) baseline achieved an AUROC of 0.83 and an AUPR of 0.51 on the test set, establishing a linear baseline for the task. The MLP, which can model nonlinear relationships between histone modifications, achieved an equivalent AUROC of 0.83 but a modestly improved AUPR of 0.60, suggesting that nonlinear feature interactions contribute primarily to improved precision and recall on the minority classes. Strikingly, the GCN substantially outperformed both baseline models, achieving an AUROC of 0.98 and an AUPR of 0.71. The improvement in AUROC from 0.83 to 0.98 indicates that the GCN is substantially better at discriminating between all classes, while the increase in AUPR from 0.51 to 0.71 demonstrates an improvement in correctly classifying the underrepresented Class 2 (X-chromosomal) and Class 3 (HAS) GA-repeats. Importantly, the LR and MLP models receive identical input features and differ only in their capacity to model nonlinear interactions. Since both models achieve the same AUROC,

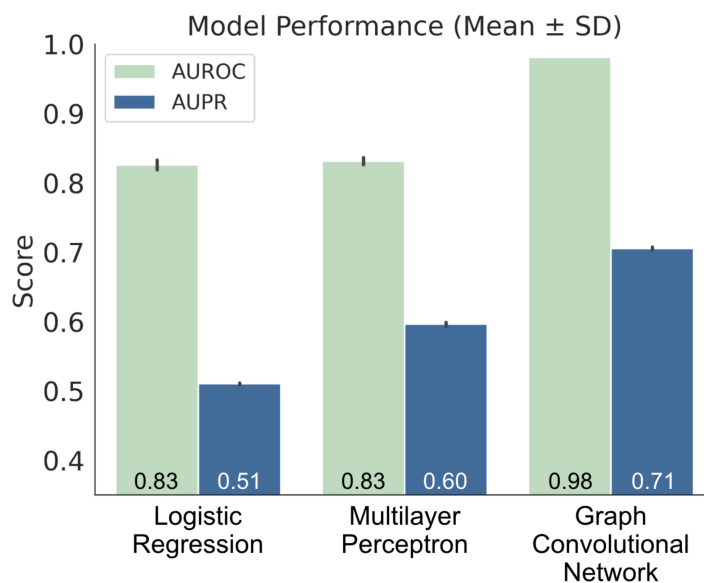


Figure 4.2: **The GCN outperforms baseline models for GA-repeat classification.** Bar plot comparing mean AUROC (green) and AUPR (blue) across logistic regression (LR), multi-layer perceptron (MLP), and graph convolutional neural network (GCN) models. Error bars represent standard deviation. The GCN achieved the highest performance on both metrics (AUROC = 0.98, AUPR = 0.71), followed by the MLP (AUROC = 0.83, AUPR = 0.60) and LR (AUROC = 0.83, AUPR = 0.51).

this suggests that linear relationships between histone modifications already capture substantial discriminative information, but that this information alone is insufficient. The dramatic improvement achieved by the GCN, which is the only model that incorporates 3D chromatin structure, demonstrates that the spatial context of histone modifications provides critical information for distinguishing between GA-repeat classes. These results support our hypothesis that 3D chromatin interaction data are essential for building robust classifiers of GA-repeat motifs.

4.4.2 Class-specific combinations of histone modifications emerge from classifier

Given that the GCN substantially outperforms the baseline models, we next sought to understand which histone modifications were most important for classifying each GA-repeat class (Fig. 4.3A). We applied GNNExplainer to derive mean normalized importance scores for each of the eight histone modification features across a random sample of 70 correctly classified test nodes per class (Fig. 4.3B). Across all three classes, the repressive mark H3K27me3 emerged as a consistently important feature, though its relative importance varied by class. For Class 1 (autosomal GA-repeats), the most important features were H3K27me3 (0.72) and H3K27ac (0.65), followed by H4K16ac (0.53) and H3K4me (0.50). The importance of the repressive mark H3K27me3

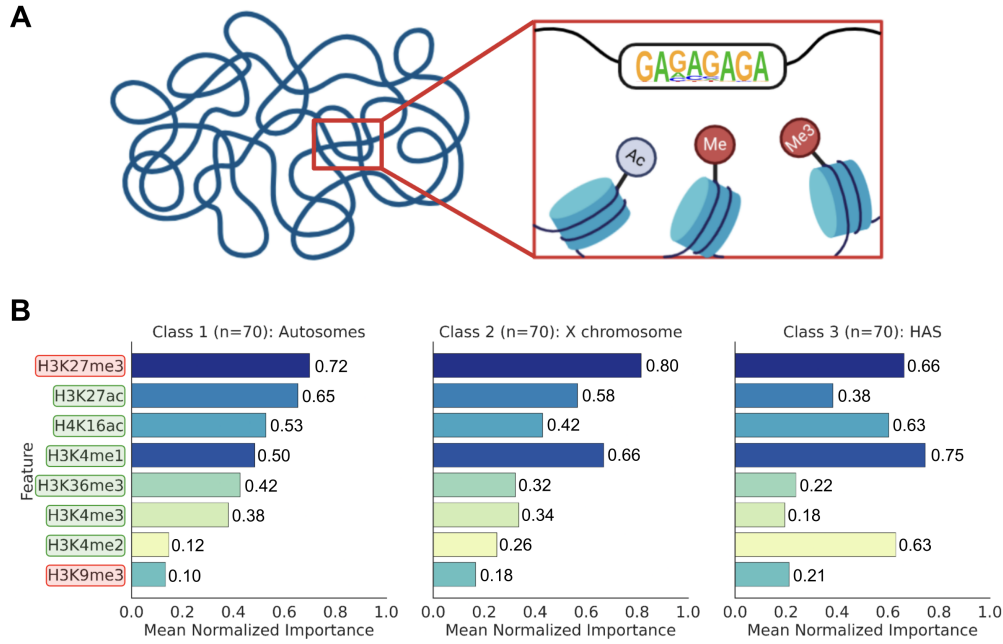


Figure 4.3: Class-specific histone modification importance scores derived from GNNExplainer. (A) Schematic illustrating the concept of the chromatin environment at a GA-repeat motif. Chromatin is organized in three-dimensional space, and the histone modifications (acetylation, methylation, trimethylation) at loci that physically contact a GA-repeat contribute to its classification. (B) Mean normalized importance scores for each of the eight histone modifications across the three GA-repeat classes (n = 70 per class). Class 1 (autosomes) shows high importance for H3K27me3, H3K27ac, and H4K16ac. Class 2 (X chromosome, non-HAS) is dominated by H3K27me3 and H3K4me1. Class 3 (HAS) shows the highest importance for H3K4me1, followed by H3K27me3, H3K4me2, and H4K16ac.

and the activating mark H3K27ac is notable as these modifications are mutually exclusive at the same residue (H3K27), suggesting that the model leverages the balance between these opposing marks to classify autosomal GA-repeats. For Class 2 (X-chromosomal, non-HAS GA-repeats), H3K27me3 was the dominant feature (0.80), followed by H3K4me1 (0.66), while H3K27ac was less important (0.58) compared to Class 1. For Class 3 (HAS GA-repeats), the importance profile was strikingly distinct: H3K4me1 was the most important feature (0.75), followed by H3K27me3 (0.66), H3K4me2 (0.63), and H4K16ac (0.63). The elevated importance of H4K16ac at HAS GA-repeats is biologically consistent with the role of this modification as the hallmark of dosage compensation, deposited by the MOF subunit of the DCC (Gelbart and Kuroda, 2009). Similarly, the high importance of H3K4me1 at HAS may reflect enhancer-associated chromatin states that facilitate DCC recruitment and spreading.

Together, these results demonstrate that each GA-repeat class possesses a distinct combinatorial pattern of histone modification importance. Furthermore, it is plausible that CLAMP and GAF have been selectively recruited to specific subsets of GA-repeat motifs based on the combinatorial histone modification landscape present in their three-dimensional chromatin neighborhood.

To further investigate this possibility, we examined the heterogeneity of epigenomic signatures at individual GA-repeat loci within each class.

4.4.3 Unique epigenome signatures are found between and within target classes

While the mean importance scores described above reveal class-level patterns, we next asked whether individual GA-repeats within the same class exhibit heterogeneous epigenomic signatures. To address this, we performed hierarchical clustering on the GNNExplainer importance scores across all sampled test nodes and histone modifications (Fig. 4). The resulting heatmap revealed that GA-repeats do not form a single homogeneous cluster per class. Instead, distinct subclusters emerge both between and within classes. At a broad level, Class 1 (autosomes) nodes largely cluster together and are characterized by high importance for H3K27ac, H3K36me3, and H3K4me3, consistent with active transcriptional states. A prominent subcluster of Class 3 (HAS) nodes (red arrow) shows elevated importance for H4K16ac and H3K27me3, suggesting a dosage compensation signature with a fine-tuned balance between active and repressive chromatin marks (Li et al. 2025). Another subcluster of Class 3 (HAS) nodes (olive arrow) are dominated by H3K27ac and H3K27me3 signals with modest H4K16ac enrichment, displaying an alternate dosage compensation signature. A cluster, which corresponds to Class 2 (X chromosome), at the bottom of the heatmap (black arrow) is solely driven by H3K27me3 and H3K27ac, completely missing H4K16ac. This indicates that even the modest enrichment of H4K16ac in three-dimensional space contributes substantially to distinguishing the HAS from other regions on the X chromosome. The presence of within-class heterogeneity suggests that GA-repeats of the same class can reside in distinct chromatin environments.

This study offers a framework to characterize the epigenomic signatures present at CLAMP and GAF binding sites throughout the genome. By enabling node-level interpretation through GNNExplainer, this approach has the potential to address key limitations of the experimental analyses presented in Chapters 2 and 3. For example, on the autosomes, where both CLAMP and GAF target contiguous GA-repeats and co-occupy many of the same loci, traditional genomic assays cannot readily distinguish the chromatin features that favor one factor over the other. A node-level analysis of importance scores at individual autosomal GA-repeats could reveal whether distinct epigenomic signatures underlie the preferential binding of CLAMP versus GAF at these shared

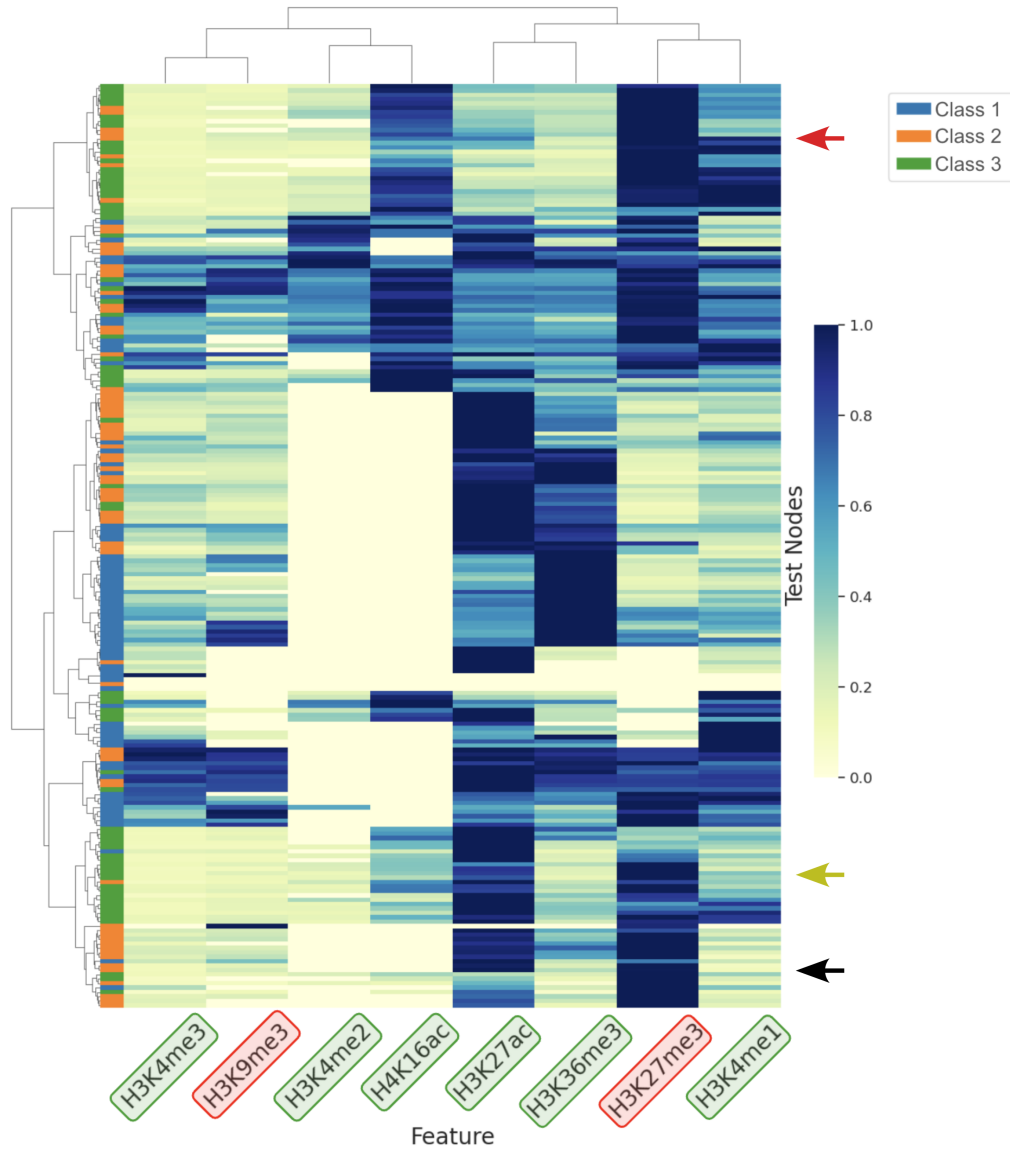


Figure 4.4: **Hierarchical clustering of GNNExplainer importance scores reveals unique epigenomic signatures within and between GA-repeat classes.** Heatmap displaying the normalized importance scores for each histone modification (columns) across individual test nodes (rows), with class identity indicated by the color bar on the left (blue = Class 1, autosomes; orange = Class 2, X chromosome; green = Class 3, HAS). Hierarchical clustering was performed on both rows (nodes) and columns (features). Red arrow indicates a subcluster of Class 3 nodes enriched for H4K16ac and H3K27me3 importance. Olive arrow indicates a subcluster of Class 3 nodes comprising elevated H3K27ac and H3K27me3 importances with modest enrichment of H4K16ac. Black arrow highlights a distinct cluster largely driven by H3K27ac and H3K27me3 features, primarily comprising Class 2 nodes.

sites. Moreover, in this study, we focused on understanding the enrichment of histone modifications present throughout the genome, leaving out edge importance. In future studies, we aim to understand which edges are the most important in determining the class of GA-repeats, which will further elucidate the spatial regulatory logic driving pioneer transcription factor binding.

Chapter 5: Discussion

Joseph Aguilera¹ and Erica Larschan¹

I wrote this chapter with editorial feedback from Dr. Erica Larschan.

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

5.1 A single cytosine within GA-repeats dictates pioneer transcription factor identity and 3D chromatin architecture

A central theme of this thesis is that minor sequence variation within transcription factor binding motifs has a substantial effect on transcription factor occupancy, chromatin looping, and gene regulation (Maurano et al. 2012). In Chapter 2, I demonstrated via *de novo* motif analysis that direct CLAMP-dependent loop anchors at high-affinity sites (HAS) on the X chromosome are distinguished from GAF-dependent anchors by a single highly conserved cytosine nucleotide that interrupts the GA-repeats (67.3% frequency in the CLAMP position weight matrix versus its absence in GAF-bound motifs) (Aguilera et al. 2026). This sequence change was known to favor CLAMP but not GAF binding biochemically (Kaye et al. 2018) but its mechanism of action *in vivo* was unknown. My Micro-C and Hi-ChIP data reveal the functional consequence of this very small change in sequence within the same type of GA-rich motif. I demonstrated that CLAMP binds preferentially to cytosine-interrupted GA-repeats and mediates short-range chromatin loops that directly link HAS to active dosage-compensated housekeeping genes (Aguilera et al. 2026). In contrast, GAF occupies uninterrupted GA-repeats and organizes a broader set of long-range interactions between transcriptionally silent insulator regions. On autosomes, GAF mediates a broad size range of chromatin loops between active regions while CLAMP also regulates short range loops between active genes. My findings provide new insight into the longstanding question of how the X chromosome is identified for dosage compensation because I determine how X-chromosome specific chromatin loops are formed that recruit the dosage compensation complex (Lucchesi and Kuroda 2015).

The implication of this finding extends beyond *Drosophila* dosage compensation. A single cytosine interruption within a GA-repeat motif can act as a molecular switch, determining which of two competing pioneer transcription factors activates gene expression. My data demonstrate that the competition between these two factors is not solely determined by differences in protein concentration or timing of expression, but rather by the precise nucleotide composition of the binding site itself (Kuzu et al. 2016). At HAS on the X chromosome, evolutionary accumulation of cytosine interruptions within GA-repeats has shifted the competitive balance decisively toward CLAMP which directly contacts the dosage compensation complex (DCC) (Soruco et al. 2013), creating an X-

specific chromatin environment capable of DCC recruitment. On autosomes, where uninterrupted GA-repeats dominate, GAF and CLAMP co-occupy sites and regulate chromatin contacts between active regions. In Chapter 2, I identified several factors that may be pursued to assist our understanding of how CLAMP or GAF are favored at autosomal loci, such as DREF which is enriched at housekeeping genes on the X chromosome and autosomes (Zabidi et al. 2015). However, an experimental, combinatorial study requires extensive time and resources to perform, underscoring the power of leveraging machine learning (Chapter 4). With unique epigenomic signatures emerging at GA-repeats throughout the genome, CLAMP and GAF may possess exclusive signatures on the autosomes which may drive a specific locus's preference for one factor over another.

This nucleotide-level regulation of transcription factor competition raises an important question: if a single cytosine can redirect the binding of a pioneer factor and restructure the 3D chromatin landscape, what happens when that cytosine is disrupted by a naturally occurring single nucleotide polymorphism (SNP)? Due to their repetitive nature, the GA-repeat motifs that CLAMP and GAF compete over are inherently susceptible to replication slippage and point mutations, making them plausible hotspots for regulatory variation (Gemayel et al. 2010; Gymrek et al. 2016). The mechanistic details of how such mutations would reshape transcription factor competition and 3D chromatin organization are explored in the following section.

5.2 Transcription factor competition at shared motifs offers a framework for interpreting non-coding genetic variation

Genome-wide association studies (GWAS) have identified thousands of single nucleotide polymorphisms (SNPs) associated with disease and complex traits (Maurano et al. 2012), yet the vast majority of these variants reside in non-coding regions of the genome whose functional significance remains unknown. Despite major efforts to document human genetic variation, we still cannot explain what most non-coding SNPs do (ENCODE Project Consortium 2012). Interestingly, many of these SNPs fall within repetitive *cis*-regulatory elements—such as enhancers, insulators, and transcription factor binding sites—where their effects on gene expression are presumed but rarely demonstrated at a mechanistic level (Horton et al. 2023). A major barrier to progress is that traditional models of transcription factor-DNA interaction treat binding as a binary event: a SNP

either abolishes a binding site, or it does not (Inukai et al. 2017). This framework is insufficient when the relevant biology involves competition between multiple factors for the same motif. In this scenario, a single nucleotide change does not eliminate binding but redirects it.

My body of work offers a mechanistic framework for understanding how non-coding SNPs can produce regulatory consequences. We have shown that CLAMP and GAF, two pioneer transcription factors with largely context-dependent and often opposing regulatory functions (repression versus activation), compete for occupancy at shared GA-repeat motifs, and that their competitive outcome is determined by subtle sequence variation within those motifs (Kaye et al. 2018; Kuzu et al. 2016). This means that a SNP at the critical cytosine position does not simply eliminate a binding site; it switches which transcription factor occupies that site and therefore which downstream regulatory program is executed. A C-to-A transversion at a CLAMP-preferred motif could convert it into a GAF-preferred site, replacing short-range loops that connect active dosage-compensated genes with broader, long-range interactions between insulator regions. Conversely, a point mutation that introduces a cytosine into a contiguous GA-repeat could create a novel CLAMP binding site where GAF previously dominated. In either case, the consequences cascade through 3D chromatin organization, altering the transcriptional output of multiple nearby genes. Notably, TF binding sites may be especially susceptible to such mutations as TF occupancy directly competes with mismatch repair machinery, elevating mutation rates at bound regulatory elements (Zhu et al. 2025).

This competitive-redirection model may help explain a category of non-coding variants that current approaches miss. Standard variant-effect prediction tools score SNPs based on whether they disrupt a known transcription factor motif (Zhou and Troyanskaya 2015), but they do not account for what happens when the disrupted motif is shared between two factors with different functions. Our data suggest that many non-coding SNPs within repetitive GA-repeat regulatory elements, which are abundant throughout eukaryotic genomes (Gemayel et al. 2010), could function not by eliminating transcription factor binding but by shifting the competitive equilibrium between co-occupying factors. Since CLAMP and GAF binding sites engage in multi-scale chromatin looping (Aguilera et al. 2026), from 1kb to inter-TAD interactions, a single SNP at one anchor point could exert pleiotropic effects on multiple genes through the reorganization of 3D chromatin interactions. I propose that systematically identifying shared binding motifs where multiple transcription factors

compete could substantially narrow the interpretation gap for non-coding disease-associated variation (Zhang et al. 2021).

An important open question that emerges from this framework is whether the regulatory impact of such SNPs is constitutive or temporally restricted. The concept of dynamic quantitative trait loci (dynamic QTLs) adds another layer of complexity (Fairfax et al. 2014; Strober et al. 2019). Some genetic variants may only exert their effects during narrow windows of development or under specific environmental conditions, remaining functionally silent at all other times. Our work provides direct evidence that this is plausible at the molecular level. In Chapter 3, we showed that CLAMP regulates 75% of heat-stress-induced transcriptional repression through the inhibition of chromatin loop formation, but this repressive function is only engaged upon heat stress (Aguilera et al. 2025). A SNP that weakens CLAMP binding at a heat-stress-responsive regulatory element would have no detectable consequence under normal growth conditions. However, under heat stress, it could impair the repressive response, leading to aberrant accumulation of toxic RNA and protein products (Richter et al. 2010). The phenotypic effect of this variant would be entirely invisible unless measured during the brief stress window.

Analogous scenarios are easy to envision in mammalian development. During brief windows of cell fate determination, pioneer factors establish enhancer accessibility at loci controlling lineage commitment (Zaret 2020). A SNP within a shared TF binding motif at one of these loci could redirect TF competition during that critical period, altering developmental gene expression in a manner which would be undetectable in adult tissues. The identification and functional characterization of such temporally restricted variants represents a hurdle in regulatory genetics, and frameworks like the one established here, where TF competition is resolved at the single-nucleotide level, will be essential for overcoming it.

5.3 A chromatin-based regulatory scaffold for heat stress response

In Chapter 3, we successfully define how the GA-binding transcription factor CLAMP represses genes after heat stress both directly and indirectly by rearranging the 3D chromatin architecture (Aguilera et al. 2025). However, we hypothesized that CLAMP may not act alone in this process and tested by probing external datasets. Prior to heat stress, we identified that CLAMP-dependent

heat-stress-repressed genes are enriched for the insulator proteins Ibf1/2, ZIPIC, CP190, and the chromatin remodeler ISWI (Cuartero et al. 2014; Maksimenko et al. 2015; Bushey et al. 2009; Deuring et al. 2000). Interestingly, we revealed that Ibf1/2 directly associates with CLAMP at a subset of repressed genes, indicating that these insulators may have a critical function post heat stress (Aguilera et al. 2025). The identification of potential direct and indirect regulatory modules (Ibf1/2, ZIPIC, CP190, and ISWI) suggests that specific protein networks may serve as a regulatory scaffold whose transcriptional output is dictated by external stimuli (Van Bortle and Corces 2012). Importantly, given that this context lacks the dosage compensation machinery necessary for activation, as described in Chapter 2, CLAMP-dependent loop anchors and repressed-genes may be primed with the necessary components to drive repression. However, within the scope of this thesis, we may only hypothesize that this is the case. We will need to perform additional experiments, ChIP-seq and possibly Hi-ChIP, to characterize the function of Ibf1/2, ZIPIC, CP190, and ISWI during the heat stress response.

The biophysical properties of the factors assembled at CLAMP-dependent loop anchors may also determine transcriptional outcomes. CLAMP is an intrinsically disordered protein with a prion-like domain (Kaye et al. 2017; Ray et al. 2024), and temperature increases are known to influence phase separation properties of disordered proteins with such domains (Rawat et al. 2021; Zhang et al. 2022; Dignon et al. 2019; Chen et al. 2022). We have recently shown that CLAMP can undergo a phase transition (Ray et al. 2024), raising the possibility that heat stress induces biophysical changes at CLAMP-associated loop anchors. Additionally, CLAMP interacts with NELF, which forms condensates under stress to drive transcriptional repression (Urban et al. 2017; Rawat et al. 2021). The presence of different combinations of chromatin-bound factors (Ibf1/2, ZIPIC, and ISWI) at different CLAMP-bound loci may regulate the phase transition properties of each locus, determining whether a gene is activated or repressed upon heat stress. To define how these factor combinations regulate phase behavior, future studies that utilize single-molecule imaging and structural approaches will be required to test this model.

In Chapter 2, I describe, in the context of male *Drosophila*, how the X chromosome has accumulated high-affinity sites that contain GA-repeats motifs which are interrupted by a single cytosine nucleotide (Kaye et al. 2018; Kuzu et al. 2016). However, regarding all other sites in the genome and in *Drosophila* females, contiguous GA-motifs may require unique chromatin environments which

can shift their competitive equilibrium to prefer one transcription factor over another (Chapter 4). Although future work lies ahead, we have successfully identified testable, candidate insulators and a chromatin modifier, which may assist CLAMP in driving repression post heat stress.

5.4 CLAMP produces opposing transcriptional outcomes in different regulatory contexts

Together, the findings from Chapters 2 and 3 establish CLAMP as a versatile pioneer transcription factor capable of both activating and repressive functions depending on context. In male *Drosophila* S2 cells, CLAMP promotes dosage compensation by organizing short-range chromatin loops that link HAS to active housekeeping genes, facilitating DCC recruitment and two-fold transcriptional upregulation (Fig. 5.1A) (Aguilera et al. 2026; Larschan et al. 2011). In female *Drosophila* Kc167 cells under heat stress, CLAMP functions as the primary mediator of transcriptional repression, inhibiting the formation of chromatin loops at gene-rich loci to suppress constitutive gene expression (Fig. 5.1B) (Aguilera et al. 2025). Notably, CLAMP occupancy at its binding sites does not change dramatically upon heat stress (Aguilera et al. 2025), indicating that the regulatory switch between activation and repression is not driven by altered CLAMP binding but by changes in the biophysical properties of the chromatin environment and the availability of context-specific co-factors (Ray et al. 2024).

This finding provides new insight into how pioneer transcription factors function beyond constitutive chromatin openers by demonstrating that they can also actively modulate downstream regulation in a context-dependent manner (Zaret 2020). Our data support a model in which CLAMP opens chromatin (Urban et al. 2017) but also modulates the 3D chromatin architecture in a context-dependent manner, such as promoting loops in one context and inhibiting them in another (Aguilera et al. 2026; Aguilera et al. 2025). The phase transition properties of CLAMP under heat stress may provide one biophysical mechanism by which identical CLAMP-DNA interactions produce different effects on loop formation across conditions (Ray et al. 2024).

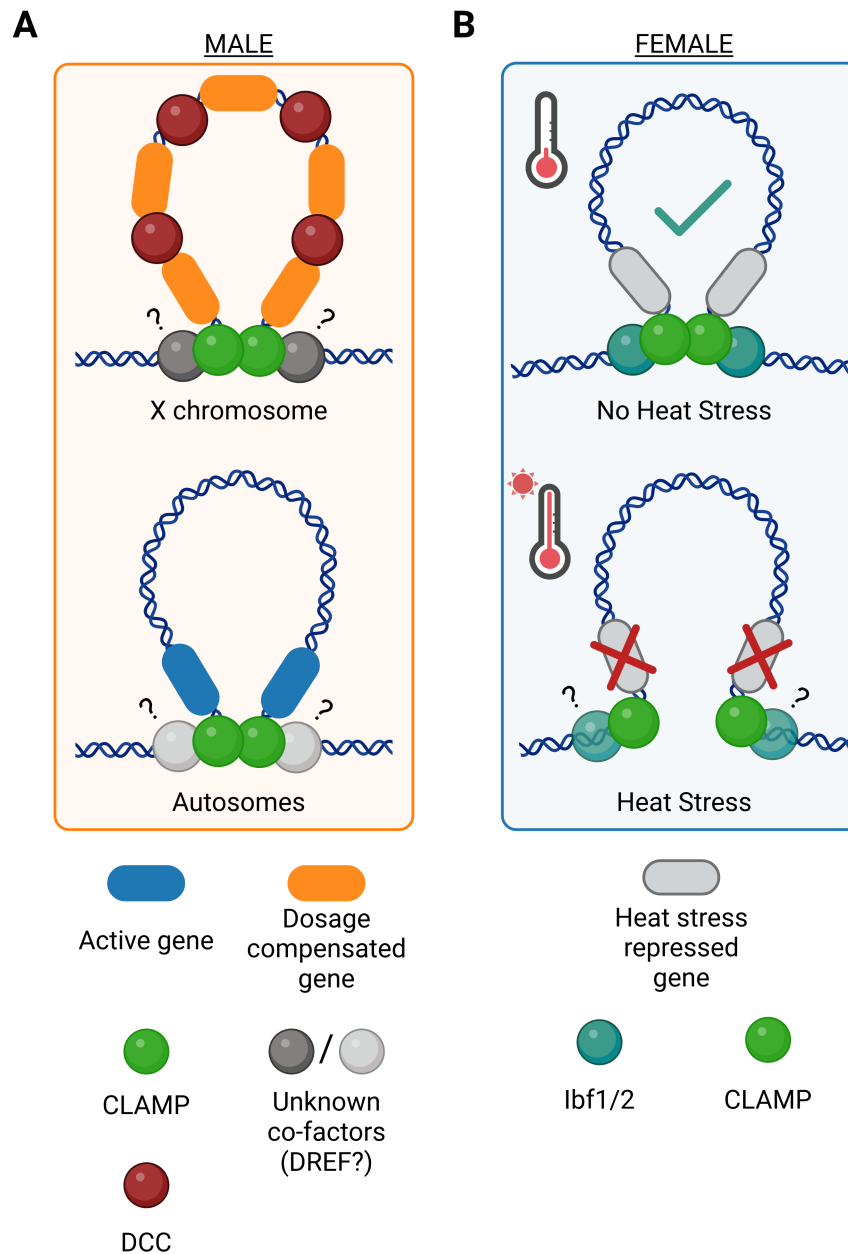


Figure 5.1: **CLAMP performs context-specific regulatory roles in male and female *Drosophila* cell lines.** (A) CLAMP upregulates nearby dosage-compensated genes on the male X chromosome by mediating 3D chromatin loops. (B) CLAMP represses nearby genes all throughout the female genome upon heat stress by preventing chromatin loop formation.

5.5 Pitfalls to consider

Several technical considerations should be taken into account when interpreting these findings. The cross-linking-based assays central to our work, such as ChIP-seq and Hi-ChIP, capture protein-DNA associations through formaldehyde fixation. Although the target protein will be enriched in the readout, nearby or co-occupying proteins may also be captured through this process. Therefore, some signal attributed to CLAMP or GAF may reflect co-captured factors, which is particularly relevant for our classification of direct versus indirect loops in Chapter 2 and the distinction between direct and indirect CLAMP-dependent HS-repressed genes in Chapter 3. Additionally, our use of RNAi-mediated knockdown rather than complete genetic knockouts means that residual CLAMP or GAF protein may partially maintain loops that would otherwise be fully lost. Unique to Chapter 3, SLAM-seq captures nascent transcription during a defined window that may not fully represent steady-state consequences over longer timescales.

Despite these limitations, the integration of multiple independent assays—such as Micro-C, Hi-ChIP, and SLAM-seq—provides strong lines of evidence that converge on the same biological model. Conclusions supported by multiple datasets represent the most robust findings, while those dependent on a single assay are interpreted cautiously.

5.6 Summary

This thesis demonstrates that the competition between CLAMP and GAF at shared GA-repeat motifs generates context-specific 3D chromatin architectures that drive fundamentally different transcriptional programs. In males, CLAMP outcompetes GAF at cytosine-interrupted GA-repeats on the X chromosome to organize local chromatin loops that recruit the dosage compensation complex. In females, under heat stress, CLAMP inhibits chromatin loop formation at gene-rich loci to mediate the repression of constitutive genes. We propose that these findings provide a mechanistic framework for understanding how single nucleotide variants at repetitive regulatory motifs and combinatorial occupancy of insulators and a chromatin modifier can produce context-dependent changes in gene expression. This body of work highlights a competitive transcription factor framework which can be used to further understand the mechanisms of dynamic QTLs in humans. Lastly,

future applications using interpretable graph neural networks have great promise in revealing the unique chromatin environments present at DNA-binding motifs all throughout the genome.

References

- Abdennur, N., & Mirny, L. A. (2020). Cooler: Scalable storage for hi-c data and other genomically labeled arrays. *Bioinformatics*, 36(1), 311–316. <https://doi.org/10.1093/bioinformatics/btz540>
- Abidi, A. A., Cattoglio, C., Tang, N. N., Fan, V. B., Dailey, G. M., Hay, A. D., Kunamaneni, P., Milkie, D. E., Darzacq, X., Betzig, E., Tjian, R., & Graham, T. G. W. (2026a). Unstructured transcription factor interactions enable emergent specificity. *Science*, eaeb6487. <https://doi.org/10.1126/science.aeb6487>
- Abidi, A. A., Cattoglio, C., Tang, N. N., Fan, V. B., Dailey, G. M., Hay, A. D., Kunamaneni, P., Milkie, D. E., Darzacq, X., Betzig, E., Tjian, R., & Graham, T. G. W. (2026b). Unstructured transcription factor interactions enable emergent specificity. *Science*, 0(0), eaeb6487. <https://doi.org/10.1126/science.aeb6487>
- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., ... Venter, J. C. (2000). The genome sequence of drosophila melanogaster. *Science*, 287(5461), 2185–2195. <https://doi.org/10.1126/science.287.5461.2185>
- Adkins, N. L., Hagerman, T. a., & Georgel, P. (2006). GAGA protein: A multi-faceted transcription factor. *Biochem. Cell Biol.*, 84(4), 559–567. <https://doi.org/10.1139/o06-062>
- Aguilera, J., Cortez, K., Segarra Alonzo, L. C., Aldana, M., Aragon, A., Gray, C., Woodman-Sousa, M., Grive, K. J., Ray, M., & Larschan, E. (2026, March 14). *Differential chromatin looping regulated by two GA-binding transcription factors creates an x-specific chromatin environment for dosage compensation* [Publication Title: bioRxiv]. <https://doi.org/10.64898/2026.03.12.711449>
- Aguilera, J., Duan, J., Cortez, K., Lee, R. S., Aragon, A., Ray, M., & Larschan, E. (2025, May 23). *The CLAMP GA-binding transcription factor regulates heat stress-induced transcriptional repression by associating with 3d chromatin loops* [Publication Title: bioRxivorg]. <https://doi.org/10.1101/2023.10.08.561401>
- Alekseyenko, A. A., Ho, J. W., Peng, S., Gelbart, M., Tolstorukov, M. Y., Plachetka, A., Kharchenko, P. V., Jung, Y. L., Gorchakov, A. A., Larschan, E., Gu, T., Minoda, A., Riddle, N. C., Schwartz, Y. B., Elgin, S. C., Karpen, G. H., Pirrotta, V., Kuroda, M. I., & Park, P. J. (2012).

- Sequence-specific targeting of dosage compensation in drosophila favors an active chromatin context. *PLoS Genet.*, 8(4), e1002646. <https://doi.org/10.1371/journal.pgen.1002646>
- Alekseyenko, A. A., McElroy, K. A., Kang, H., Zee, B. M., Kharchenko, P. V., & Kuroda, M. I. (2015). BioTAP-XL: Cross-linking/tandem affinity purification to study DNA targets, RNA, and protein components of chromatin-associated complexes. *Curr. Protoc. Mol. Biol.*, 109(1), 21.30.1–21.30.32. <https://doi.org/10.1002/0471142727.mb2130s109>
- Alekseyenko, A. A., Peng, S., Larschan, E., Gorchakov, A. A., Lee, O.-K., Kharchenko, P., McGrath, S. D., Wang, C. I., Mardis, E. R., Park, P. J., & Kuroda, M. I. (2008). A sequence motif within chromatin entry sites directs MSL establishment on the drosophila x chromosome. *Cell*, 134(4), 599–609. <https://doi.org/10.1016/j.cell.2008.06.033>
- Alipanahi, B., DeLong, A., Weirauch, M. T., & Frey, B. J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.*, 33(8), 831–838. <https://doi.org/10.1038/nbt.3300>
- Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J. R., Grabska-Barwinska, A., Taylor, K. R., Assael, Y., Jumper, J., Kohli, P., & Kelley, D. R. (2021, April 8). *Effective gene expression prediction from sequence by integrating long-range interactions* [Publication Title: bioRxiv]. <https://doi.org/10.1101/2021.04.07.438649>
- Ay, F., Bailey, T. L., & Noble, W. S. (2014). Statistical confidence estimation for hi-c data reveals regulatory chromatin contacts. *Genome Res.*, 24(6), 999–1011. <https://doi.org/10.1101/gr.160374.113>
- Bag, I., Dale, R., Palmer, C., & Lei, E. (2019). The zinc-finger protein CLAMP promotes gypsy chromatin insulator function in drosophila. *Journal of cell science*, 132, jcs226092.
- Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.*, 21(3), 381–395. <https://doi.org/10.1038/cr.2011.22>
- Barth, R., Janissen, R., Muras, L., van der Torre, J., Litos, G., van der Sluis, E., van den Berg, A., Davidson, I. F., Peters, J.-M., & Dekker, C. (2025). Two CTCF motifs impede cohesin-mediated DNA loop extrusion. *Mol. Cell*, 85(23), 4304–4317.e9. <https://doi.org/10.1016/j.molcel.2025.11.001>
- Bhattacharyya, S., Chandra, V., Vijayanand, P., & Ay, F. (2019). Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nature communications*, 10, 4221.

- Bigness, J., Loinaz, X., Patel, S., Larschan, E., & Singh, R. (2022). Integrating long-range regulatory interactions to predict gene expression using graph convolutional networks. *J. Comput. Biol.*, 29(5), 409–424. <https://doi.org/10.1089/cmb.2021.0316>
- Boettiger, A. N., Bintu, B., Moffitt, J. R., Wang, S., Beliveau, B. J., Fudenberg, G., Imakaev, M., Mirny, L. A., Wu, C.-T., & Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, 529(7586), 418–422. <https://doi.org/10.1038/nature16496>
- Boija, A., Klein, I. A., Sabari, B. R., Dall’Agnese, A., Coffey, E. L., Zamudio, A. V., Li, C. H., Shrinivas, K., Manteiga, J. C., Hannett, N. M., Abraham, B. J., Afeyan, L. K., Guo, Y. E., Rimel, J. K., Fant, C. B., Schuijers, J., Lee, T. I., Taatjes, D. J., & Young, R. A. (2018). Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell*, 175(7), 1842–1855.e16. <https://doi.org/10.1016/j.cell.2018.10.042>
- Bonev, B., & Cavalli, G. (2016). Organization and function of the 3d genome. *Nat. Rev. Genet.* https://idp.nature.com/authorize/casa?redirect_uri=https://www.nature.com/articles/nrg.2016.112&casa_token=hrCeJTpB_FQAAAAA:B3mS_tWyNZOLDrYrkLDoySmWdGGEywhMg8-NobzzPsVBy4Cr5AluojK-pJSqc27G4uCLkN6mTtmomVQEg
- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, & Greenleaf WJ. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*, 10, 1213–1218. <https://www.pubchase.com/article/24097267>
- Bushey, A. M., Ramos, E., & Corces, V. G. (2009). Three subclasses of a drosophila insulator show distinct and cell type-specific genomic distributions. *Genes Dev.*, 23(11), 1338–1350. <https://doi.org/10.1101/gad.1798209>
- Chen, D., Lyu, M., Kou, X., Li, J., Yang, Z., Gao, L., Li, Y., Fan, L.-m., Shi, H., & Zhong, S. (2022). Integration of light and temperature sensing by liquid-liquid phase separation of phytochrome b. *Molecular Cell*, 82, 3015–3029.
- Chen, Y., Lun, A., & Smyth, G. (2016). From reads to genes to pathways: Differential expression analysis of RNA-seq experiments using rsubread and the edgeR quasi-likelihood pipeline. *F1000Research*, 5, 1438.
- Choppakatla, P., Patel, A. L., Borjigin, T., Udomlumleart, T., Hu, J., Gregor, T., Boettiger, A. N., & Levine, M. S. (2026, February 18). *Loop extrusion accelerates long-range enhancer-*

- promoter searches in living embryos* [Publication Title: bioRxivorg]. <https://doi.org/10.64898/2026.02.17.706355>
- Chowdhary, S., Kainth, A. S., Paracha, S., Gross, D. S., & Pincus, D. (2022). Inducible transcriptional condensates drive 3d genome reorganization in the heat shock response. *Molecular Cell*, 82(22), 4386–4399.e7. <https://doi.org/10.1016/j.molcel.2022.10.013>
- Chu, W.-T., & Wang, J. (2021). Thermodynamic and sequential characteristics of phase separation and droplet formation for an intrinsically disordered region/protein ensemble. *PLoS computational biology*, 17, e1008672.
- Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M., & Zaret, K. S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell*, 9(2), 279–289. [https://doi.org/10.1016/s1097-2765\(02\)00459-8](https://doi.org/10.1016/s1097-2765(02)00459-8)
- Clapier, C. R., Iwasa, J., Cairns, B. R., & Peterson, C. L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.*, 18(7), 407–422. <https://doi.org/10.1038/nrm.2017.26>
- Coactivator condensation at super-enhancers links phase separation and gene control | science.* (n.d.). Retrieved April 3, 2026, from <https://www.science.org/doi/10.1126/science.aar3958>
- Core, L. J., Waterfall, J. J., & Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, 322(5909), 1845–1848. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19056941
- Cuartero, S., Fresán, U., Reina, O., Planet, E., & Espinàs, M. (2014). Ibf1 and Ibf2 are novel CP190-interacting proteins required for insulator function. *The EMBO journal*, 33, 637–647.
- Cubenas-Potts, C., Rowley, M., Lyu, X., Li, G., Lei, E., & Corces, V. (2017). Different enhancer classes in drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3d architecture. *Nucleic acids research*, 45, 1714–1730.
- Cugusi, S., Mitter, R., Kelly, G., Walker, J., Han, Z., Pisano, P., Wierer, M., Stewart, A., & Svejstrup, J. (2022). Heat shock induces premature transcript termination and reconfigures the human transcriptome. *Molecular Cell*, 82, 1573–1588.
- Dale, R. K., Pedersen, B. S., & Quinlan, A. R. (2011). Pybedtools: A flexible python library for manipulating genomic datasets and annotations. *Bioinformatics*, 27(24), 3423–3424. <https://doi.org/10.1093/bioinformatics/btr539>

- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, 10(2). <https://doi.org/10.1093/gigascience/giab008>
- Dann, G., Liszczak, G., Bagert, J., Müller, M., Nguyen, U., Wojcik, F., Brown, Z., Bos, J., Panchenko, T., & Pihl, R. (2017). ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference. *Nature*, 548, 607–611.
- Dekker, J., & Mirny, L. (2016). The 3d genome as moderator of chromosomal communication. *Cell*, 164(6), 1110–1121. <https://doi.org/10.1016/j.cell.2016.02.007>
- Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science*, 295(5558), 1306–1311. <https://doi.org/10.1126/science.1067799>
- Deng, X., Berletch, J. B., Ma, W., Nguyen, D. K., Hiatt, J. B., Noble, W. S., Shendure, J., & Disteche, C. M. (2013). Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. *Dev. Cell*, 25(1), 55–68. <https://doi.org/10.1016/j.devcel.2013.01.028>
- Deuring, R., Fanti, L., Armstrong, J. A., Sarte, M., Papoulas, O., Prestel, M., Daubresse, G., Verardo, M., Moseley, S. L., Berloco, M., Tsukiyama, T., Wu, C., Pimpinelli, S., & Tamkun, J. W. (2000). The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Mol. Cell*, 5(2), 355–365. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10882076
- Dignon, G. L., Zheng, W., Kim, Y. C., & Mittal, J. (2019). Temperature-controlled liquid-liquid phase separation of disordered proteins. *ACS Cent. Sci.*, 5(5), 821–830. <https://doi.org/10.1021/acscentsci.9b00102>
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398), 376–380. <https://doi.org/10.1038/nature11082>
- Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., Rubio, E. D., Krumm, A., Lamb, J., Nusbaum, C., Green, R. D., & Dekker, J. (2006). Chromosome conformation capture carbon copy (5c): A massively parallel solution for mapping interactions between genomic elements. *Genome Res.*, 16(10), 1299–1309. <https://doi.org/10.1101/gr.5571506>

- Drouin, J. (2014). Minireview: Pioneer transcription factors in cell fate specification. *Mol. Endocrinol.*, 28(7), 989–998. <https://doi.org/10.1210/me.2014-1084>
- Duan, J., Rieder, L., Colonna, M., Huang, A., Mckenney, M., Watters, S., Deshpande, G., Jordan, W., Fawzi, N., & Larschan, E. (2021). CLAMP and zelda function together to promote drosophila zygotic genome activation. *Elife*, 10, e69937.
- Duarte, F., Fuda, N., Mahat, D., Core, L., Guertin, M., & Lis, J. (2016). Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. *Genes & development*, 30, 1731–1746.
- Dynamic regulation of endogenous transcription factor hubs at single-molecule resolution | bioRxiv.* (n.d.). Retrieved April 3, 2026, from <https://www.biorxiv.org/content/10.64898/2026.01.25.701606v1.abstract>
- Eggers, N., & Becker, P. B. (2021). Cell-free genomics reveal intrinsic, cooperative and competitive determinants of chromatin interactions. *Nucleic Acids Res.*, 49(13), 7602–7617. <https://doi.org/10.1093/nar/gkab558>
- Eggers, N., Gkountromichos, F., Krause, S., Campos-Sparr, A., & Becker, P. B. (2023, April 11). *Physical interaction between MSL2 and CLAMP assures direct co-operativity and prevents competition at composite binding sites* [Publication Title: bioRxiv]. <https://doi.org/10.1101/2023.04.11.536365>
- ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74. <https://doi.org/10.1038/nature11247>
- The essential drosophila CLAMP protein differentially regulates non-coding roX RNA expression and dosage compensation-mediated transcription.* (n.d.).
- Ewels, P., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M., Di Tommaso, P., & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nature biotechnology*, 38, 276–278.
- Fairfax, B. P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., & Knight, J. C. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*, 343(6175), 1246949. <https://doi.org/10.1126/science.1246949>
- Fey, M., & Lenssen, J. E. (2019, March 6). *Fast graph representation learning with PyTorch geometric* [ISBN: 1903.02428 Publication Title: arXiv [cs.LG]]. <https://doi.org/10.48550/arXiv.1903.02428>

- Flyamer, I. M., Illingworth, R. S., & Bickmore, W. A. (2020). Coolpup.py: Versatile pile-up analysis of hi-c data. *Bioinformatics*, 36(10), 2980–2985. <https://doi.org/10.1093/bioinformatics/btaa073>
- Freund, M. M., Harrison, M. M., & Torres-Zelada, E. F. (2024). Exploring the reciprocity between pioneer factors and development. *Development*, 151(13). <https://doi.org/10.1242/dev.201921>
- Fuda, N. J., Guertin, M. J., Sharma, S., Danko, C. G., Martins, A. L., Siepel, A., & Lis, J. T. (2015). GAGA factor maintains nucleosome-free regions and has a role in RNA polymerase II recruitment to promoters. *PLoS Genet.*, 11(3). <https://doi.org/10.1371/journal.pgen.1005108>
- Gaskill, M. M., Gibson, T. J., Larson, E. D., & Harrison, M. M. (2021). GAF is essential for zygotic genome activation and chromatin accessibility in the early drosophila embryo. *Elife*, 10. <https://doi.org/10.7554/eLife.66668>
- Gelbart, M. E., & Kuroda, M. I. (2009). Drosophila dosage compensation: A complex voyage to the x chromosome. *Development*, 136(9), 1399–1410. <https://doi.org/10.1242/dev.029645>
- Gemayel, R., Vences, M. D., Legendre, M., & Verstrepen, K. J. (2010). Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu. Rev. Genet.*, 44(1), 445–477. <https://doi.org/10.1146/annurev-genet-072610-155046>
- Gerstein, M. B., Lu, Z. J., Van Nostrand, E. L., Cheng, C., Arshinoff, B. I., Liu, T., Yip, K. Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., Alves, P., Chateigner, A., Perry, M., Morris, M., Auerbach, R. K., Feng, X., Leng, J., Vielle, A., Niu, W., ... Waterston, R. H. (2010). Integrative analysis of the caenorhabditis elegans genome by the modENCODE project. *Science*, 330(6012), 1775–1787. <https://doi.org/10.1126/science.1196914>
- Goenka, A., Sengupta, S., Pandey, R., Parihar, R., Mohanta, G., Mukerji, M., & Ganesh, S. (2016). Human satellite-III non-coding RNAs modulate heat-shock-induced transcriptional repression. *Journal of cell science*, 129, 3541–3552.
- Gordân, R., Shen, N., Dror, I., Zhou, T., Horton, J., Rohs, R., & Bulyk, M. L. (2013). Genomic regions flanking e-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep.*, 3(4), 1093–1104. <https://doi.org/10.1016/j.celrep.2013.03.014>
- Gramates, L. S., Agapite, J., Attrill, H., Calvi, B. R., Crosby, M. A., Dos Santos, G., Goodman, J. L., Goutte-Gattat, D., Jenkins, V. K., Kaufman, T., Larkin, A., Matthews, B. B., Millburn,

- G., Strelets, V. B., & the FlyBase Consortium. (2022). FlyBase: A guided tour of highlighted features. *Genetics*, 220(4). <https://doi.org/10.1093/genetics/iyac035>
- Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., Wildonger, J., & O'Connor-Giles, K. M. (2013). Genome engineering of drosophila with the CRISPR RNA-guided cas9 nuclease. *Genetics*, 194(4), 1029–1035. <https://doi.org/10.1534/genetics.113.152710>
- Greenberg, A. J., Yanowitz, J. L., & Schedl, P. (2004). The drosophila GAGA factor is required for dosage compensation in males and for the formation of the male-specific-lethal complex chromatin entry site at 12de. *Genetics*, 166(1), 279–289. <https://doi.org/10.1534/genetics.166.1.279>
- Gu, L., Reilly, P. F., Lewis, J. J., Reed, R. D., Andolfatto, P., & Walters, J. R. (2019). Dichotomy of dosage compensation along the neo z chromosome of the monarch butterfly. *Curr. Biol.*, 29(23), 4071–4077.e3. <https://doi.org/10.1016/j.cub.2019.09.056>
- Gu, W., Szauter, P., & Lucchesi, J. C. (1998). Targeting of MOF, a putative histone acetyl transferase, to the x chromosome of drosophila melanogaster. *Dev. Genet.*, 22(1), 56–64. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9499580
- Guertin, M., Petesch, S., Zobeck, K., Min, I., & Lis, J. (2010). Cold spring harbor symposia on quantitative biology. *Cold Spring Harbor Laboratory Press*, 75, 1–9.
- Guertin, M., & Lis, J. (2010). Chromatin landscape dictates HSF binding to target DNA elements. *PLoS genetics*, 6, e1001114.
- Gutierrez-Perez, I., Rowley, M. J., Lyu, X., Valadez-Graham, V., Vallejo, D. M., Ballesta-Illan, E., Lopez-Atalaya, J. P., Kremsky, I., Caparros, E., Corces, V. G., & Dominguez, M. (2019). Ecdysone-induced 3d chromatin reorganization involves active enhancers bound by pip-squeak and polycomb. *Cell Rep.*, 28(10), 2715–2727.e5. <https://doi.org/10.1016/j.celrep.2019.07.096>
- Gymrek, M., Willems, T., Guilmatre, A., Zeng, H., Markus, B., Georgiev, S., Daly, M. J., Price, A. L., Pritchard, J. K., Sharp, A. J., & Erlich, Y. (2016). Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat. Genet.*, 48(1), 22–29. <https://doi.org/10.1038/ng.3461>

- Hamilton, W. L., Ying, R., & Leskovec, J. (2017, June 7). *Inductive representation learning on large graphs* [ISBN: 1706.02216 Publication Title: arXiv [cs.SI]]. <https://doi.org/10.48550/arXiv.1706.02216>
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Mol. Cell*, *38*(4), 576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>
- Herzog, V., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T., Wlotzka, W., Von Haeseler, A., Zuber, J., & Ameres, S. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. *Nature methods*, *14*, 1198–1204.
- Hochheimer, A., Zhou, S., Zheng, S., Holmes, M. C., & Tjian, R. (2002). TRF2 associates with DREF and directs promoter-selective gene expression in drosophila. *Nature*, *420*(6914), 439–445. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12459787
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.*, *9*(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Horton, C. A., Alexandari, A. M., Hayes, M. G. B., Marklund, E., Schaepe, J. M., Aditham, A. K., Shah, N., Suzuki, P. H., Shrikumar, A., Afek, A., Greenleaf, W. J., Gordân, R., Zeitlinger, J., Kundaje, A., & Fordyce, P. M. (2023). Short tandem repeats bind transcription factors to tune eukaryotic gene expression. *Science*, *381*(6664), eadd1250. <https://doi.org/10.1126/science.add1250>
- Hsieh, T., Cattoglio, C., Slobodyanyuk, E., Hansen, A., Rando, O., Tjian, R., & Darzacq, X. (2020). Resolving the 3d landscape of transcription-linked mammalian chromatin folding. *Mol Cell*, *78*, 539–553.
- Hsieh, T.-H. S., Fudenberg, G., Goloborodko, A., & Rando, O. J. (2016). Micro-c XL: Assaying chromosome conformation from the nucleosome to the entire genome. *Nat. Methods*, *13*(12), 1009–1011. <https://doi.org/10.1038/nmeth.4025>
- Hsieh, T.-H. S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., & Rando, O. J. (2015). Mapping nucleosome resolution chromosome folding in yeast by micro-c. *Cell*, *162*(1), 108–119. <https://doi.org/10.1016/j.cell.2015.05.048>
- Huang, Y., An, J., Sircar, S., Bergis, C., Lopes, C., He, X., Da Costa, B., Tan, F.-Q., Bazin, J., & Antunez-Sanchez, J. (2023). HSFA1a modulates plant heat stress responses and alters

- the 3d chromatin organization of enhancer-promoter interactions. *Nature communications*, 14, 469.
- Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., ... Morgan, M. (2015). Orchestrating high-throughput genomic analysis with bioconductor. *Nat. Methods*, 12(2), 115–121. <https://doi.org/10.1038/nmeth.3252>
- Hutin, S., Kumita, J., Strotmann, V., Dolata, A., Ling, W., Louafi, N., Popov, A., Milhiet, P.-E., Blackledge, M., & Nanao, M. (2023). Phase separation and molecular ordering of the prion-like domain of the thermosensory protein EARLY FLOWERING 3. *bioRxiv*, 2023.2003.2012.532276.
- Infantino, S., Jones, S. A., Walker, J. A., Maxwell, M. J., Light, A., O'Donnell, K., Tsantikos, E., Peperzak, V., Phesse, T., Ernst, M., Mackay, F., Hibbs, M. L., Fairfax, K. A., & Tarlinton, D. M. (2014). The tyrosine kinase lyn limits the cytokine responsiveness of plasma cells to restrict their accumulation in mice. *Sci. Signal.*, 7(338), ra77. <https://doi.org/10.1126/scisignal.2005105>
- Inukai, S., Kock, K. H., & Bulyk, M. L. (2017). Transcription factor–DNA binding: Beyond binding site motifs. *Current Opinion in Genetics and Development*, 43, 110–119. <https://doi.org/10.1016/j.gde.2017.02.007>
- Ito, T., Bulger, M., Pazin, M. J., Kobayashi, R., & Kadonaga, J. T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell*, 90(1), 145–155. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9230310
- Iwafuchi-Doi, M., & Zaret, K. S. (2014). Pioneer transcription factors in cell reprogramming. *Genes Dev.*, 28(24), 2679–2692. <https://doi.org/10.1101/gad.253443.114>
- Jaeger, S. A., Chan, E. T., Berger, M. F., Stottmann, R., Hughes, T. R., & Bulyk, M. L. (2010). Conservation and regulatory associations of a wide affinity range of mouse transcription factor binding sites. *Genomics*, 95(4), 185–195. <https://doi.org/10.1016/j.ygeno.2010.01.002>
- Jeong, R., & Bulyk, M. L. (2025). Meta-analysis reveals transcription factors and DNA binding domain variants associated with congenital heart defect and orofacial cleft. *NPJ Genom. Med.*, 10(1), 63. <https://doi.org/10.1038/s41525-025-00525-0>

- Jiang, N., Emberly, E., Cuvier, O., & Hart, C. M. (2009). Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in drosophila melanogaster links BEAF to transcription. *Mol. Cell. Biol.*, 29(13), 3556–3568. <https://doi.org/10.1128/MCB.01748-08>
- Jordan, W., & Larschan, E. (2021). The zinc finger protein CLAMP promotes long-range chromatin interactions that mediate dosage compensation of the drosophila male x-chromosome. *Epigenetics & Chromatin*, 14, 1–16.
- Kahn, T., Savitsky, M., Kuong, C., Jacquier, C., Cavalli, G., Chang, J.-M., & Schwartz, Y. (2023). Topological screen identifies hundreds of cp190-and CTCF-dependent drosophila chromatin insulator elements. *Science Advances*, 9, eade0090.
- Karbalayghareh, A., Sahin, M., & Leslie, C. S. (2022). Chromatin interaction-aware gene regulatory modeling with graph attention networks. *Genome Res.*, 32(5), 930–944. <https://doi.org/10.1101/gr.275870.121>
- Kaul, A., Bhattacharyya, S., & Ay, F. (2020). Identifying statistically significant chromatin contacts from hi-c data with FitHiC2. *Nat. Protoc.*, 15(3), 991–1012. <https://doi.org/10.1038/s41596-019-0273-0>
- Kaushal, A., Dorier, J., Wang, B., Mohana, G., Taschner, M., Cousin, P., Waridel, P., Iseli, C., Semenova, A., & Restrepo, S. (2022). Essential role of cp190 in physical and regulatory boundary formation. *Science Advances*, 8, eabl8834.
- Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K., & Henikoff, S. (2019). CUT&tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.*, 10(1), 1930. <https://doi.org/10.1038/s41467-019-09982-5>
- Kaye, E. G., Booker, M., Kurland, J. V., Conicella, A. E., Fawzi, N. L., Bulyk, M. L., Tolstorukov, M. Y., & Larschan, E. (2018). Differential occupancy of two GA-binding proteins promotes targeting of the drosophila dosage compensation complex to the male x chromosome. *Cell Rep.*, 22(12), 3227–3239. <https://doi.org/10.1016/j.celrep.2018.02.098>
- Kaye, E. G., Kurbidaeva, A., Wolle, D., Aoki, T., Schedl, P., & Larschan, E. (2017). Drosophila dosage compensation loci associate with a boundary-forming insulator complex. *Mol. Cell. Biol.*, 37(21). <https://doi.org/10.1128/MCB.00253-17>
- Kelley, D. R., Reshef, Y. A., Bileschi, M., Belanger, D., McLean, C. Y., & Snoek, J. (2018). Sequential regulatory activity prediction across chromosomes with convolutional neural networks. *Genome Res.*, 28(5), 739–750. <https://doi.org/10.1101/gr.227819.117>

- Kelley, R. L., Meller, V. H., Gordadze, P. R., Roman, G., Davis, R. L., & Kuroda, M. I. (1999). Epigenetic spreading of the drosophila dosage compensation complex from roX RNA genes into flanking chromatin. *Cell*, 98(4), 513–522. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10481915
- Kellner, W., Ramos, E., Van Bortle, K., Takenaka, N., & Corces, V. (2012). Genome-wide phosphoacetylation of histone h3 at drosophila enhancers and promoters. *Genome research*, 22, 1081–1088.
- Kharchenko, P. V., Alekseyenko, A. a., Schwartz, Y. B., Minoda, A., Riddle, N. C., Ernst, J., Sabo, P. J., Larschan, E., Gorchakov, A. a., Gu, T., Linder-Basso, D., Plachetka, A., Shanower, G., Tolstorukov, M. Y., Luquette, L. J., Xi, R., Jung, Y. L., Park, R. W., Bishop, E. P., ... Park, P. J. (2010). Comprehensive analysis of the chromatin landscape in drosophila melanogaster. *Nature*. <https://doi.org/10.1038/nature09725>
- Khetan, S., Carroll, B. S., & Bulyk, M. L. (2025). Multiple overlapping binding sites determine transcription factor occupancy. *Nature*, 646(8086), 1001–1011. <https://doi.org/10.1038/s41586-025-09472-3>
- Kipf, T. N., & Welling, M. (2016, September 9). *Semi-supervised classification with graph convolutional networks* [ISBN: 1609.02907 Publication Title: arXiv [cs.LG]]. <https://doi.org/10.48550/arXiv.1609.02907>
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, 128(4), 693–705. <https://doi.org/10.1016/j.cell.2007.02.005>
- Kuzu, G., Kaye, E. G., Chery, J., Siggers, T., Yang, L., Dobson, J. R., Boor, S., Bliss, J., Liu, W., Jogl, G., Rohs, R., Singh, N. D., Bulyk, M. L., Tolstorukov, M. Y., & Larschan, E. (2016). Expansion of GA dinucleotide repeats increases the density of CLAMP binding sites on the x-chromosome to promote drosophila dosage compensation. *PLoS Genet.*, 12(7). <https://doi.org/10.1371/journal.pgen.1006120>
- Kwak, H., Fuda, N. J., Core, L. J., & Lis, J. T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science*, 339(6122), 950–953. <https://doi.org/10.1126/science.1229386>
- Kyrchanova, O., & Georgiev, P. (2014). Chromatin insulators and long-distance interactions in drosophila. *FEBS Letters*, 588(1), 8–14. <https://doi.org/10.1016/j.febslet.2013.10.039>

- Lakhotia, S., Mallik, M., Singh, A., & Ray, M. (2012). The large noncoding hsrw-n transcripts are essential for thermotolerance and remobilization of hnRNPs, HP1 and RNA polymerase II during recovery from heat shock in drosophila. *Chromosoma*, 121, 49–70.
- Larschan, E., Soruco, M., Lee, O.-K., Peng, S., Bishop, E., Chery, J., Goebel, K., Feng, J., Park, P., & Kuroda, M. (2012). Identification of chromatin-associated regulators of MSL complex targeting in drosophila dosage compensation. *PLoS genetics*, 8, e1002830.
- Larschan, E., Alekseyenko, A. a., Gortchakov, A. a., Peng, S., Li, B., Yang, P., Workman, J. L., Park, P. J., & Kuroda, M. I. (2007). MSL complex is attracted to genes marked by h3k36 trimethylation using a sequence-independent mechanism. *Mol. Cell*, 28(1), 121–133. <https://doi.org/10.1016/j.molcel.2007.08.011>
- 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. <https://www.pubchase.com/article/21368835>
- Larson, E. D., Komori, H., Gibson, T. J., Ostgaard, C. M., Hamm, D. C., Schnell, J. M., Lee, C.-Y., & Harrison, M. M. (2021). Cell-type-specific chromatin occupancy by the pioneer factor zelda drives key developmental transitions in drosophila. *Nat. Commun.*, 12(1), 7153. <https://doi.org/10.1038/s41467-021-27506-y>
- Larson, E. D., Marsh, A. J., & Harrison, M. M. (2021). Pioneering the developmental frontier. *Mol. Cell*, 81(8), 1640–1650. <https://doi.org/10.1016/j.molcel.2021.02.020>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lee, T. I., & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell*, 152(6), 1237–1251. <https://doi.org/10.1016/j.cell.2013.02.014>
- Levo, M., Raimundo, J., Bing, X. Y., Sisco, Z., Batut, P. J., Ryabichko, S., Gregor, T., & Levine, M. S. (2022). Transcriptional coupling of distant regulatory genes in living embryos. *Nature*, 605(7911), 754–760. <https://doi.org/10.1038/s41586-022-04680-7>
- Li, H. (2013, March 16). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM* [ISBN: 1303.3997 Publication Title: arXiv [q-bio.GN]]. <http://arxiv.org/abs/1303.3997>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>

- Li, J., Liu, Y., Rhee, H., Ghosh, S., Bai, L., Pugh, B., & Gilmour, D. (2013). Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. *Molecular cell*, 50, 711–722.
- Li, J., & Gilmour, D. S. (2013). Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and m1bp, a novel transcription factor. *EMBO J.*, 32(13), 1829–1841. <https://doi.org/10.1038/emboj.2013.111>
- Li, J., Xu, S., Liu, Z., Yang, L., Ming, Z., Zhang, R., Zhao, W., Peng, H., Quinn, J. J., Wu, M., Geng, Y., Zhang, Y., He, J., Chen, M., Li, N., Shao, N.-Y., & Ma, Q. (2025). A noncanonical role of roX RNAs in autosomal epigenetic repression. *Nat. Commun.*, 16(1), 155. <https://doi.org/10.1038/s41467-024-55711-y>
- Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H., Ong, C.-T., Cubeñas-Potts, C., Hu, M., Lei, E., & Bosco, G. (2015). Widespread rearrangement of 3d chromatin organization underlies polycomb-mediated stress-induced silencing. *Molecular cell*, 58, 216–231.
- Li, X., Tang, X., Bing, X., Catalano, C., Li, T., Dolsten, G., Wu, C., & Levine, M. (2023). GAGA-associated factor fosters loop formation in the drosophila genome. *Molecular cell*, 83, 1519–1526.
- Liang, Z., Zhang, Q., Ji, C., Hu, G., Zhang, P., Wang, Y., Yang, L., & Gu, X. (2021). Reorganization of the 3d chromatin architecture of rice genomes during heat stress. *BMC biology*, 19, 1–10.
- Lieberman-aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., & Mirny, L. A. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(9), 289–293. <https://doi.org/10.1038/nature08398>
- Lomaev, D., Mikhailova, A., Erokhin, M., Shaposhnikov, A. V., Moresco, J. J., Blokhina, T., Wolle, D., Aoki, T., Ryabykh, V., Yates, J. R., Shidlovskii, Y. V., Georgiev, P., Schedl, P., & Chetverina, D. (2017). The GAGA factor regulatory network: Identification of GAGA factor associated proteins. *PLoS One*, 12(3), 1–20. <https://doi.org/10.1371/journal.pone.0173602>
- Lucchesi, J. C., & Kuroda, M. I. (2015). Dosage compensation in drosophila. *Cold Spring Harb. Perspect. Biol.*, 7(5). <https://doi.org/10.1101/cshperspect.a019398>

- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648), 251–260. <https://doi.org/10.1038/38444>
- Mahat, D., Salamanca, H., Duarte, F., Danko, C., & Lis, J. (2016). Mammalian heat shock response and mechanisms underlying its genome-wide transcriptional regulation. *Molecular cell*, 62, 63–78.
- Maksimenko, O., Bartkuhn, M., Stakhov, V., Herold, M., Zolotarev, N., Jox, T., Buxa, M., Kirsch, R., Bonchuk, A., & Fedotova, A. (2015). Two new insulator proteins, pita and ZIPIC, target CP190 to chromatin. *Genome research*, 25, 89–99.
- Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutayavin, T., Stehling-Sun, S., Johnson, A. K., Canfield, T. K., Giste, E., Diegel, M., ... Stamatoyannopoulos, J. A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science*, 337(6099), 1190–1195. <https://doi.org/10.1126/science.1222794>
- Maziak, N., Zhang, Y., Groll, F., Brown, H. E., Madich, A., Kaur, Y., Harrison, M. M., Zhou, J., & Vaquerizas, J. M. (2026). Three-dimensional genome reorganization foreshadows zygotic genome activation in drosophila. *Nat. Genet.*, 58(3), 607–617. <https://doi.org/10.1038/s41588-026-02503-3>
- Meller, V. H., & Rattner, B. P. (2002). The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *EMBO J.*, 21(5), 1084–1091. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11867536
- Melnikova, L., Juge, F., Gruzdeva, N., Mazur, A., Cavalli, G., & Georgiev, P. (2004). Interaction between the GAGA factor and mod(mdg4) proteins promotes insulator bypass in drosophila. *Proc. Natl. Acad. Sci. U. S. A.*, 101(41), 14806–14811. <https://doi.org/10.1073/pnas.0403959101>
- Merkenschlager, M., & Nora, E. P. (2016). CTCF and cohesin in genome folding and transcriptional gene regulation. *Annu. Rev. Genomics Hum. Genet.*, 17, 17–43. <https://doi.org/10.1146/annurev-genom-083115-022339>
- Merkenschlager, M., & Odom, D. T. (2013). CTCF and cohesin: Linking gene regulatory elements with their targets. *Cell*, 152(6), 1285–1297. <https://doi.org/10.1016/j.cell.2013.02.029>

- modENCODE Consortium, Roy, S., Ernst, J., Kharchenko, P. V., Kheradpour, P., Negre, N., Eaton, M. L., Landolin, J. M., Bristow, C. A., Ma, L., Lin, M. F., Washietl, S., Arshinoff, B. I., Ay, F., Meyer, P. E., Robine, N., Washington, N. L., Di Stefano, L., Berezikov, E., ... Kellis, M. (2010). Identification of functional elements and regulatory circuits by drosophila modENCODE. *Science*, 330(6012), 1787–1797. <https://doi.org/10.1126/science.1198374>
- Mohajan, S., & Gross, D. S. (2025). Transcriptional condensates and the nuclear pore complex regulate gene expression and 3d genome architecture in response to stress. *Biochemical Society Transactions*, 53(5), 1295–1309. <https://doi.org/10.1042/BST20253086>
- Mumbach, M., Rubin, A., Flynn, R., Dai, C., Khavari, P., Greenleaf, W., & Chang, H. (2016). HiChIP: Efficient and sensitive analysis of protein-directed genome architecture. *Nature methods*, 13, 919–922.
- Neumann, T., Herzog, V., Muhar, M., von Haeseler, A., Zuber, J., Ameres, S., & Rescheneder, P. (2019). Quantification of experimentally induced nucleotide conversions in high-throughput sequencing datasets. *BMC bioinformatics*, 20, 1–16.
- Nora, E. P., Goloborodko, A., Valton, A.-L., Gibcus, J. H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L. A., & Bruneau, B. G. (2017). Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*, 169(5), 930–944.e22. <https://doi.org/10.1016/j.cell.2017.05.004>
- Ogiyama, Y., Schuettengruber, B., Papadopoulos, G., Chang, J.-M., & Cavalli, G. (2018). Polycomb-dependent chromatin looping contributes to gene silencing during drosophila development. *Molecular cell*, 71, 73–88.
- Open2C, Abdennur, N., Fudenberg, G., Flyamer, I. M., Galitsyna, A. A., Goloborodko, A., Imakaev, M., & Venev, S. V. (2024). Pairtools: From sequencing data to chromosome contacts. *PLoS Comput. Biol.*, 20(5), e1012164. <https://doi.org/10.1371/journal.pcbi.1012164>
- Park, P. J. (2009). ChIP-seq: Advantages and challenges of a maturing technology. *Nat. Rev. Genet.*, 10(10), 669–680. <https://doi.org/10.1038/nrg2641>
- Park, Y., Kelley, R. L., Oh, H., Kuroda, M. I., & Meller, V. H. (2002). Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins. *Science*, 298(5598), 1620–1623. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12446910
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chil-

- amkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019, December 3). *PyTorch: An imperative style, high-performance deep learning library* [ISBN: 1912.01703 Publication Title: arXiv [cs.LG]]. <https://doi.org/10.48550/arXiv.1912.01703>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2012, January 2). *Scikit-learn: Machine learning in python* [ISBN: 1201.0490 Publication Title: arXiv [cs.LG]]. <https://doi.org/10.48550/arXiv.1201.0490>
- Perrimon, N., Ni, J.-Q., & Perkins, L. (2010). In vivo RNAi: Today and tomorrow. *Cold Spring Harb. Perspect. Biol.*, 2(8), a003640. <https://doi.org/10.1101/cshperspect.a003640>
- Poliacikova, G., Barthez, M., Rival, T., Aouane, A., Luis, N. M., Richard, F., Daian, F., Brouilly, N., Schnorrer, F., Maurel-Zaffran, C., Graba, Y., & Saurin, A. J. (2023). M1bp is an essential transcriptional activator of oxidative metabolism during drosophila development. *Nat. Commun.*, 14(1), 3187. <https://doi.org/10.1038/s41467-023-38986-5>
- Prayitno, K., Schauer, T., Regnard, C., & Becker, P. B. (2019). Progressive dosage compensation during drosophila embryogenesis is reflected by gene arrangement. *EMBO Rep.*, 20(8), e48138. <https://doi.org/10.15252/embr.201948138>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K. C., Grüning, B. A., Villaveces, J., Habermann, B., Akhtar, A., & Manke, T. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.*, 9(1), 189. <https://doi.org/10.1038/s41467-017-02525-w>
- Ramírez, F., Lingg, T., Toscano, S., Lam, K. C., Georgiev, P., Chung, H.-R., Lajoie, B. R., de Wit, E., Zhan, Y., de Laat, W., Dekker, J., Manke, T., & Akhtar, A. (2015). High-affinity sites form an interaction network to facilitate spreading of the MSL complex across the x chromosome in drosophila. *Mol. Cell*, 60(1), 146–162. <https://doi.org/10.1016/j.molcel.2015.08.024>
- Rawat, P., Boehning, M., Hummel, B., Aprile-Garcia, F., Pandit, A., Eisenhardt, N., Khavaran, A., Niskanen, E., Vos, S., & Palvimo, J. (2021). Stress-induced nuclear condensation of NELF drives transcriptional downregulation. *Molecular cell*, 81, 1013–1026.

- Ray, J., Munn, P., Vihervaara, A., Lewis, J., Ozer, A., Danko, C., & Lis, J. (2019). Chromatin conformation remains stable upon extensive transcriptional changes driven by heat shock. *Proceedings of the National Academy of Sciences*, 116, 19431–19439.
- Ray, M., Zaborowsky, J., Mahableshwarkar, P., Vaidyanathan, S., Shum, J., Viswanathan, R., Huang, A., Wang, S.-H., Johnson, V., & Wake, N. (2024). Dual DNA/RNA-binding factor regulates dynamics of hnRNP splicing condensates. *bioRxiv*, 2024.2001.2011.575216.
- Richter, K., Haslbeck, M., & Buchner, J. (2010). The heat shock response: Life on the verge of death. *Molecular cell*, 40, 253–266.
- Rieder, L. E., Koreski, K. P., Boltz, K. A., Kuzu, G., Urban, J. A., Bowman, S., Zeidman, A., William T. Jordan, I. I. I., Tolstorukov, M. Y., Marzluff, W. F., Duronio, R. J., & Larschan, E. N. (2017). Histone locus regulation by the drosophila dosage compensation adaptor protein CLAMP. *Genes Dev.* <https://doi.org/10.1101/gad.300855.117>
- Rieder, L. E., Jordan, W. T., 3rd, & Larschan, E. N. (2019). Targeting of the dosage-compensated male x-chromosome during early drosophila development. *Cell Rep.*, 29(13), 4268–4275.e2. <https://doi.org/10.1016/j.celrep.2019.11.095>
- Robinson, M., McCarthy, D., & Smyth, G. (2010). edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *bioinformatics*, 26, 139–140.
- Rowley, M., Nichols, M., Lyu, X., Ando-Kuri, M., Rivera, I., Hermetz, K., Wang, P., Ruan, Y., & Corces, V. (2017). Evolutionarily conserved principles predict 3d chromatin organization. *Molecular cell*, 67, 837–852.
- Samata, M., & Akhtar, A. (2018). Dosage compensation of the x chromosome: A complex epigenetic assignment involving chromatin regulators and long noncoding RNAs. *Annu. Rev. Biochem.*, 87, 323–350. <https://doi.org/10.1146/annurev-biochem-062917-011816>
- Schones, D. E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G., & Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. *Cell*, 132(5), 887–898. <https://doi.org/10.1016/j.cell.2008.02.022>
- Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., & Barillot, E. (2015). HiC-pro: An optimized and flexible pipeline for hi-c data processing. *Genome Biol.*, 16(1), 259. <https://doi.org/10.1186/s13059-015-0831-x>
- Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., & de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered

- by chromosome conformation capture-on-chip (4c). *Nat. Genet.*, 38(11), 1348–1354. <https://doi.org/10.1038/ng1896>
- Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife*, 6. <https://doi.org/10.7554/eLife.21856>
- Skok, Tsirigos, Mazzoni, & Reinberg. (n.d.). CRISPR and biochemical screens identify MAZ as a cofactor in CTCF-mediated insulation at hox clusters. *Nature*. <https://www.nature.com/articles/s41588-021-01008-5>
- Soruco, M., Chery, J., Bishop, E., Siggers, T., Tolstorukov, M., Leydon, A., Sugden, A., Goebel, K., Feng, J., & Xia, P. (2013). The CLAMP protein links the MSL complex to the x chromosome during drosophila dosage compensation. *Genes & development*, 27, 1551–1556.
- Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.*, 13(9), 613–626. <https://doi.org/10.1038/nrg3207>
- Stansfield, J. C., Cresswell, K. G., Vladimirov, V. I., & Dozmorov, M. G. (2018). HiCcompare: An r-package for joint normalization and comparison of Hi-c datasets. *BMC Bioinformatics*, 19(1). <https://doi.org/10.1186/s12859-018-2288-x>
- Straub, T., Grimaud, C., Gilfillan, G. D., Mitterweger, A., & Becker, P. B. (2008). The chromosomal high-affinity binding sites for the drosophila dosage compensation complex. *PLoS Genet.*, 4(12), e1000302. <https://doi.org/10.1371/journal.pgen.1000302>
- Strober, B. J., Elorbany, R., Rhodes, K., Krishnan, N., Tayeb, K., Battle, A., & Gilad, Y. (2019). Dynamic genetic regulation of gene expression during cellular differentiation. *Science*, 364(6447), 1287–1290. <https://doi.org/10.1126/science.aaw0040>
- Sun, L., Jing, Y., Liu, X., Li, Q., Xue, Z., Cheng, Z., Wang, D., He, H., & Qian, W. (2020). Heat stress-induced transposon activation correlates with 3d chromatin organization rearrangement in arabidopsis. *Nature communications*, 11, 1886.
- Teves, S., & Henikoff, S. (2011). Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. *Genes & development*, 25, 2387–2397.
- Tikhonova, E., Fedotova, A., Bonchuk, A., Mogila, V., Larschan, E. N., Georgiev, P., & Maksimenko, O. (2019). The simultaneous interaction of MSL2 with CLAMP and DNA provides redundancy in the initiation of dosage compensation in drosophila males. *Development*, 146(19). <https://doi.org/10.1242/dev.179663>
- Tiusanen, V., Patel, D., Xia, J., Xu, C., Datta, S., Yun, J., Cancellieri, S., Fei, L., Yilmaz, M., Pitkänen, E., Prekovic, S., Pihlajamaa, P., & Sahu, B. (2026, January 23). Classical enhancers

- couple cis-regulatory logic with transcriptional condensates and 3d genome architecture [ISSN: 2692-8205 Pages: 2026.01.23.701252 Section: New Results]. <https://doi.org/10.64898/2026.01.23.701252>
- Tsukiyama, T., Daniel, C., Tamkun, J., & Wu, C. (1995). ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell*, 83(6), 1021–1026. [https://doi.org/10.1016/0092-8674\(95\)90217-1](https://doi.org/10.1016/0092-8674(95)90217-1)
- Urban, J., Kuzu, G., Bowman, S., Scruggs, B., Henriques, T., Kingston, R., Adelman, K., Tolstorukov, M., & Larschan, E. (2017). Enhanced chromatin accessibility of the dosage compensated drosophila male x-chromosome requires the CLAMP zinc finger protein. *PloS one*, 12, e0186855.
- Urban, J., Urban, J., Kuzu, G., & Larschan, E. (2017). The drosophila CLAMP protein associates with diverse proteins on chromatin. *PloS one*, 12, e0189772.
- Valsecchi, C. I. K., Basilicata, M. F., Georgiev, P., Gaub, A., Seyfferth, J., Kulkarni, T., Panhale, A., Semplicio, G., Manjunath, V., Holz, H., Dasmeh, P., & Akhtar, A. (2020). RNA nucleation by MSL2 induces selective x chromosome compartmentalization. *Nature*, 589(7840), 137–142. <https://doi.org/10.1038/s41586-020-2935-z>
- Valsecchi, C. I. K., Basilicata, M. F., Semplicio, G., Georgiev, P., Gutierrez, N. M., & Akhtar, A. (2018). Facultative dosage compensation of developmental genes on autosomes in drosophila and mouse embryonic stem cells. *Nat. Commun.*, 9(1), 3626. <https://doi.org/10.1038/s41467-018-05642-2>
- Van Bortle, K., Nichols, M., Li, L., Ong, C.-T., Takenaka, N., Qin, Z., & Corces, V. (2014). Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome biology*, 15, 1–18.
- Van Bortle, K., Ramos, E., Takenaka, N., Yang, J., Wahi, J., & Corces, V. (2012). Drosophila CTCF tandemly aligns with other insulator proteins at the borders of h3k27me3 domains. *Genome research*, 22, 2176–2187.
- Van Bortle, K., & Corces, V. G. (2012). Nuclear organization and genome function. *Annu. Rev. Cell Dev. Biol.*, 28(1), 163–187. <https://doi.org/10.1146/annurev-cellbio-101011-155824>
- Varisco, M., Cavalheiro, G. R., Rodriguez Viales, R., Schaub, C., Girardot, C., Forneris, M., Comeault, A. A., & Furlong, E. E. M. (2025, September 6). *Heterotypic directional motifs contribute to TAD boundary function in drosophila* [Publication Title: bioRxiv]. <https://doi.org/10.1101/2025.09.06.674533>

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017, June 12). *Attention is all you need* [ISBN: 1706.03762 Publication Title: arXiv [cs.CL]]. <http://arxiv.org/abs/1706.03762>
- Vian, L., Pękowska, A., Rao, S. S. P., Kieffer-Kwon, K.-R., Jung, S., Baranello, L., Huang, S.-C., El Khattabi, L., Dose, M., Pruett, N., Sanborn, A. L., Canela, A., Maman, Y., Oksanen, A., Resch, W., Li, X., Lee, B., Kovalchuk, A. L., Tang, Z., ... Casellas, R. (2018). The energetics and physiological impact of cohesin extrusion. *Cell*, 175(1), 292–294. <https://doi.org/10.1016/j.cell.2018.09.002>
- Vihervaara, A., Duarte, F. M., & Lis, J. T. (2018). Molecular mechanisms driving transcriptional stress responses. *Nat. Rev. Genet.*, 19(6), 385–397. <https://doi.org/10.1038/s41576-018-0001-6>
- Vo Ngoc, L., Kassavetis, G. A., & Kadonaga, J. T. (2019). The RNA polymerase II core promoter in drosophila. *Genetics*, 212(1), 13–24. <https://doi.org/10.1534/genetics.119.302021>
- Wallace, E., Kear-Scott, J., Pilipenko, E., Schwartz, M., Laskowski, P., Rojek, A., Katanski, C., Riback, J., Dion, M., & Franks, A. (2015). Reversible, specific, active aggregates of endogenous proteins assemble upon heat stress. *Cell*, 162, 1286–1298.
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, 10(1), 57–63. <https://doi.org/10.1038/nrg2484>
- Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., Manke, T., Backofen, R., Ramírez, F., & Grüning, B. A. (2018). Galaxy HiCExplorer: A web server for reproducible hi-c data analysis, quality control and visualization. *Nucleic Acids Res.*, 46, W11–W16. <https://doi.org/10.1093/nar/gky504>
- Wolff, J., Rabbani, L., Gilsbach, R., Richard, G., Manke, T., Backofen, R., & Grüning, B. A. (2020). Galaxy HiCExplorer 3: A web server for reproducible hi-c, capture hi-c and single-cell hi-c data analysis, quality control and visualization. *Nucleic Acids Res.*, 48, W177–W184. <https://doi.org/10.1093/nar/gkaa220>
- Wolle, D., Cleard, F., Aoki, T., Deshpande, G., Schedl, P., & Karch, F. (2015). Functional requirements for fab-7 boundary activity in the bithorax. *Mol. Cell. Biol.*, 35(21), 3739–3752. <https://doi.org/10.1128/MCB.00456-15>.Address
- Wood, A., Van Bortle, K., Ramos, E., Takenaka, N., Rohrbaugh, M., Jones, B., Jones, K., & Corces, V. (2011). Regulation of chromatin organization and inducible gene expression by a drosophila insulator. *Molecular cell*, 44, 29–38.

- Wu, C. (1995). Heat shock transcription factors: Structure and regulation. *Annu. Rev. Cell Dev. Biol.*, 11, 441–469. <https://doi.org/10.1146/annurev.cb.11.110195.002301>
- Wu, C.-H., Yamaguchi, Y., Benjamin, L., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., & Gilmour, D. (2003). NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in drosophila. *Genes & development*, 17, 1402–1414.
- Xu, W., Zhong, Q., Lin, D., Zuo, Y., Dai, J., Li, G., & Cao, G. (2021). CoolBox: A flexible toolkit for visual analysis of genomics data. *BMC bioinformatics*, 22, 1–9.
- Yan, J., Qiu, Y., Ribeiro Dos Santos, A. M., Yin, Y., Li, Y. E., Vinckier, N., Nariai, N., Benaglio, P., Raman, A., Li, X., Fan, S., Chiou, J., Chen, F., Frazer, K. A., Gaulton, K. J., Sander, M., Taipale, J., & Ren, B. (2021). Systematic analysis of binding of transcription factors to noncoding variants. *Nature*, 591(7848), 147–151. <https://doi.org/10.1038/s41586-021-03211-0>
- Yang, J., Sung, E., Donlin-Asp, P., & Corces, V. (2013). A subset of drosophila myc sites remain associated with mitotic chromosomes colocalized with insulator proteins. *Nature communications*, 4, 1464.
- Ying, R., Bourgeois, D., You, J., Zitnik, M., & Leskovec, J. (2019, March 9). *GNNExplainer: Generating explanations for graph neural networks* [ISBN: 1903.03894 Publication Title: arXiv [cs.LG]]. <https://doi.org/10.48550/arXiv.1903.03894>
- Zabidi, M. A., Arnold, C. D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., & Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*, 518(7540), 556–559. <https://doi.org/10.1038/nature13994>
- Zaret, K. S. (2020). Pioneer transcription factors initiating gene network changes. *Annu. Rev. Genet.*, 54, 367–385. <https://doi.org/10.1146/annurev-genet-030220-015007>
- Zaret, K. S., & Carroll, J. S. (2011). Pioneer transcription factors: Establishing competence for gene expression. *Genes Dev.*, 25(21), 2227–2241. <https://doi.org/10.1101/gad.176826.111>
- Zhang, H., Shao, S., Zeng, Y., Wang, X., Qin, Y., Ren, Q., Xiang, S., Wang, Y., Xiao, J., & Sun, Y. (2022). Reversible phase separation of HSF1 is required for an acute transcriptional response during heat shock. *Nature cell biology*, 24, 340–352.
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biol.*, 9(9), R137. <https://doi.org/10.1186/gb-2008-9-9-r137>

- Zhang, Y., Ho, T. D., Buchler, N. E., & Gordân, R. (2021). Competition for DNA binding between paralogous transcription factors determines their genomic occupancy and regulatory functions. *Genome Res.*, 31(7), 1216–1229. <https://doi.org/10.1101/gr.275145.120>
- Zhimulev, I. F., Belyaeva, E. S., Semeshin, V. F., Koryakov, D. E., Demakov, S. A., Demakova, O. V., Pokholkova, G. V., & Andreyeva, E. N. (2004). Polytene chromosomes: 70 years of genetic research. *Int. Rev. Cytol.*, 241, 203–275. [https://doi.org/10.1016/S0074-7696\(04\)41004-3](https://doi.org/10.1016/S0074-7696(04)41004-3)
- Zhou, J., & Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods*, 12(10), 931–934. <https://doi.org/10.1038/nmeth.3547>
- Zhu, W., Zhang, Y., Sahay, H., Wasserman, H., Afek, A., Williams, J., Shaltz, S., Johnson, C., Pinheiro, K., MacAlpine, D. M., Weninger, K. R., Erie, D. A., Jinks-Robertson, S., & Gordân, R. (2025). DNA mutagenesis driven by transcription factor competition with mismatch repair. *Cell*, 188(20), 5735–5747.e15. <https://doi.org/10.1016/j.cell.2025.07.003>
- Zimmer, F., Harrison, P. W., Dessimoz, C., & Mank, J. E. (2016). Compensation of dosage-sensitive genes on the chicken z chromosome. *Genome Biol. Evol.*, 8(4), 1233–1242. <https://doi.org/10.1093/gbe/evw075>
- Zimmer, F., Fox, A. M., Pan, Q., Rühle, F., Andersen, P., Huylmans, A.-K., Schwander, T., Basilicata, M. F., & Keller Valsecchi, C. I. (2025). Convergent evolution of h4k16ac-mediated dosage compensation in the ZW species *artemia franciscana*. *PLoS Genet.*, 21(10), e1011895. <https://doi.org/10.1371/journal.pgen.1011895>
- Zolotarev, N., Fedotova, A., Kyrchanova, O., Bonchuk, A., Penin, A., Lando, A., Eliseeva, I., Kulakovskiy, I., Maksimenko, O., & Georgiev, P. (2016). Architectural proteins pita, zw5, and ZIPIC contain homodimerization domain and support specific long-range interactions in *drosophila*. *Nucleic acids research*, 44, 7228–7241.