2017-05-15

Three-dimensional Folding of Eukaryotic Genomes

Tsung-Han S. Hsieh
University of Massachusetts Medical School

THREE-DIMENSIONAL FOLDING OF EUKARYOTIC GENOMES

A Dissertation Presented

By

TSUNG-HAN HSIEH

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 15th, 2017

Interdisciplinary Graduate Program

THREE-DIMENSIONAL FOLDING OF EUKARYOTIC GENOMES

A Dissertation Presented
By
TSUNG-HAN HSIEH

The signatures of the Dissertation Defense Committee signify
completion and approval as to style and content of the Dissertation

_____
Oliver J. Rando, M.D., Ph.D., Thesis Advisor

_____
Job Dekker, Ph.D., Member of Committee

_____
Thomas G. Fazzio, Ph.D., Member of Committee

_____
Manuel Garber, Ph.D., Member of Committee

_____
Craig L. Peterson, Ph.D., Member of Committee

_____
Fred M. Winston, Ph.D., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the
requirements of the Dissertation Committee

_____
Craig L. Peterson, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies
that the student has met all graduation requirements of the school.

_____
Anthony Carruthers, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

Interdisciplinary Graduate Program

May 15th, 2017

DEDICATION

I dedicate the dissertation to my beloved wife, best friend, and colleague (!),

Tiffany,

whose love, patience, and faith have been the essence of my soul in the past decade. She equally contributes to all my success and honors.

ACKNOWLEDGEMENTS

I would like to acknowledge my thesis advisor, Oliver Rando, who always inspires me as my research is stranded at a bottleneck, motivates me as nothing is working, guides me as I almost lost direction, and encourages me even though some ideas sound silly. I appreciate that he allows me freely chase my favorite topics and manage my work-life balance without any pressure. It was the best days of doing science in my life. I am so glad having a chance to work with the magical science dude.

I thank all the Rando lab members, past and present, who make the corner of 9F a pleasant working environment. In particular, I would like to thank Marta, Amanda, and Hsiuyi, who taught me everything since my Day 1 in the Rando lab, Jeremy, Ben, David and Markus, who made the bench works so much fun, Upasna, Caitlin, and Nils, who never mind to discuss any topic with me, and Lina, who takes care everything in the lab. I would also like to thank Melissa, who helps edit and improve my writing.

I thank the member of my TRAC committee for suggestions and support throughout my entire graduate study. Finally, I thank my collaborators Assaf, Geoff, and Nir with whom I am able to make my studies possible.

ABSTRACT

Chromatin packages eukaryotic genomes via a hierarchical series of folding steps, encrypting multiple layers of epigenetic information, which are capable of regulating nuclear transactions in response to complex signals in environment. Besides the 1-dimensinal chromatin landscape such as nucleosome positioning and histone modifications, little is known about the secondary chromatin structures and their functional consequences related to transcriptional regulation and DNA replication. The family of chromosomal conformation capture (3C) assays has revolutionized our understanding of large-scale chromosome folding with the ability to measure relative interaction probability between genomic loci *in vivo*. However, the suboptimal resolution of the typical 3C techniques leaves the levels of nucleosome interactions or 30 nm structures inaccessible, and also restricts their applicability to study gene level of chromatin folding in small genome organisms such as yeasts, worm, and plants. To uncover the “blind spot” of chromatin organization, I developed an innovative method called Micro-C and an improved protocol, Micro-C XL, which enable to map chromatin structures at all range of scale from single nucleosome to the entire genome. Several fine-scale aspects of chromatin folding in budding and fission yeasts have been identified by Micro-C, including histone tail-mediated tri-/tetra-nucleosome stackings, gene crumples/globules, and chromosomally-interacting domains

(CIDs). CIDs are spatially demarcated by the boundaries, which are colocalized with the promoters of actively transcribed genes and histone marks for active transcription or turnover. The levels of chromatin compaction are regulated via transcription-dependent or transcription-independent manner – either the perturbations of transcription or the mutations of chromatin regulators strongly affect the global chromatin folding. Taken together, Micro-C further reveals chromatin folding behaviors below the sub-kilobase scale and opens an avenue to study chromatin organization in many biological systems.

Table of Contents

LIST OF FIGURES

LIST OF COPYRIGHTED MATERIALS PRODUCED BY THE AUTHOR

- CHAPTER II is reprinted and adapted from the co-authored work with permission from Elsevier.

  License #: 4087991457238

  Hsieh, T.H., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., and Rando, O.J. (2015). Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. Cell 162, 108-119.

- CHAPTER III is reprinted and adapted from the co-authored work with permission from Nature Publishing Group.

  License #: permission is not required for the authors.

  Hsieh, T.H., Fudenberg, G., Goloborodko, A., and Rando, O.J. (2016). Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. Nat Methods 13, 1009-1011.

- CHAPTER IV is reprinted and adapted from the co-authored work with permission from Elsevier.

  License #: permission is not required for the authors.

  Weiner, A.*, Hsieh, T.H.*, Appleboim, A.*, Chen, H.V., Rahat, A., Amit, I., Rando, O.J., and Friedman, N. (2015). High-resolution chromatin dynamics during a yeast stress response. Mol Cell 58, 371-386.

LIST OF THIRD PARTY COPYRIGHTED MATERIAL

Figure #	Publisher	License #
Figure 1.1	American Society for Microbiology	Open resource
	Nature Publishing Group	4087701073648
	Nature Publishing Group	4087701499071
Figure 1.2	Nature Publishing Group	4087710063839
Figure 1.3	American Society for Microbiology	Open resource
Figure 1.4	InTech	Open resource
Figure 1.5	Elsevier	4087710448664
Figure 1.6	Nature Publishing Group	4098940294131
	AAAS	4087710608681
	ASBMB	Open resource
Figure 1.7	Springer	Open resource
Figure 1.8	Nature Publishing Group	4087710195152
	Figshare	Personal contact
Figure 1.9	Nature Publishing Group	4087710195152
Figure 1.10	Nature Publishing Group	4087701365567

PREFACE

- The work in CHAPTER II was a collaborative project between the Rando, Friedman, and Dekker labs, in which experiments were designed by THH and OJR and performed by THH, data was developed and analyzed by THH, AW and OJR, the manuscript was prepared by THH, AW, NF, and OJR. Early data analysis was guided by BL and JD. This has been published as:

  Hsieh, T.H., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., and Rando, O.J. (2015). Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. Cell 162, 108-119.

- The work in CHAPTER III was a collaborative project between the Rando and Mirny labs. Experiments were designed by THH and OJR and performed by THH. Data was analyzed by THH, GF, AG, and OJR. The manuscript was prepared by THH, GF, and OJR. This has been published as:

  Hsieh, T.H., Fudenberg, G., Goloborodko, A., and Rando, O.J. (2016). Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. Nat Methods 13, 1009-1011.

- The work in CHAPTER IV was a collaborative project between the Rando and Friedman labs. Experiments were designed by THH, AW, NF, and OJR and performed by THH and AW. Data was analyzed by AW, AA, NF, and OJR. The manuscript was prepared by AW, THH, AA, NF, and OJR. Early experiments were guided by HVC. This has been published as:

  Weiner, A.*, Hsieh, T.H.*, Appleboim, A.*, Chen, H.V., Rahat, A., Amit, I., Rando, O.J., and Friedman, N. (2015). High-resolution chromatin dynamics during a yeast stress response. Mol Cell 58, 371-386.

- The preliminary works in CHAPTER V were initiated and designed by THH and OJR. THH performed experiments and analyzed data.

CHAPTER I

Hierarchical Architecture of the Eukaryotic Genome: From Textbook to Modern Era

Overview

Chromatin packages eukaryotic genomes via a hierarchical series of folding steps (**Figure 1.1**), encrypting multiple layers of genetic and epigenetic information such as histone modifications and chromatin organization, etc., which are capable of regulating nuclear transactions in response to complex signals in environment. In eukaryotes, 147 base-pairs of DNA are first wrapped around histones (H3/H4 and H2A/H2B dimers) into nucleosomes. The first level of genome compaction is organization into the characteristic “beads on a string” structure, with nucleosomes separated by relatively accessible linker DNA. Our understanding of this primary structure of chromatin is well-developed, with multiple crystal structures solved for the nucleosome, and a plethora of genome-wide studies that identify the positions and the covalent modifications of individual nucleosomes across the genome in various organisms, in some cases at single nucleotide-resolution. The next step in chromosome folding remains relatively poorly-characterized – for example, the long-held belief that chromatin fibers form a helical secondary structure termed the 30 nm fiber was readily observed under *in vitro* conditions, but its existence *in vivo* was the subject of intense debate. At the higher-order scale, the genome is further organized as

multiple-megabase of euchromatin and heterochromatin, which is then further condensed as mitotic chromosomes during the cell cycle. Although traditional biochemical and microscopic studies of the past have developed our fundamental knowledge of chromosome organization, a few key features of chromatin are still missing. How does chromatin dynamically respond to drastic change in environment? More specifically, how do histone modifications react to a massive transcriptional reprogramming? How are nucleosomal arrays folded *in vivo*? What kind of genomic information do these multiple levels of chromatin structures encode? This dissertation will discuss what we have learned about chromatin structures from the classic aspect to the modern view, and what the problems are unsolved (**Chapter I, Introduction**), and how a newly developed technique fills in the gaps of knowledge in chromatin (**Chapter II, III, IV, Results**).

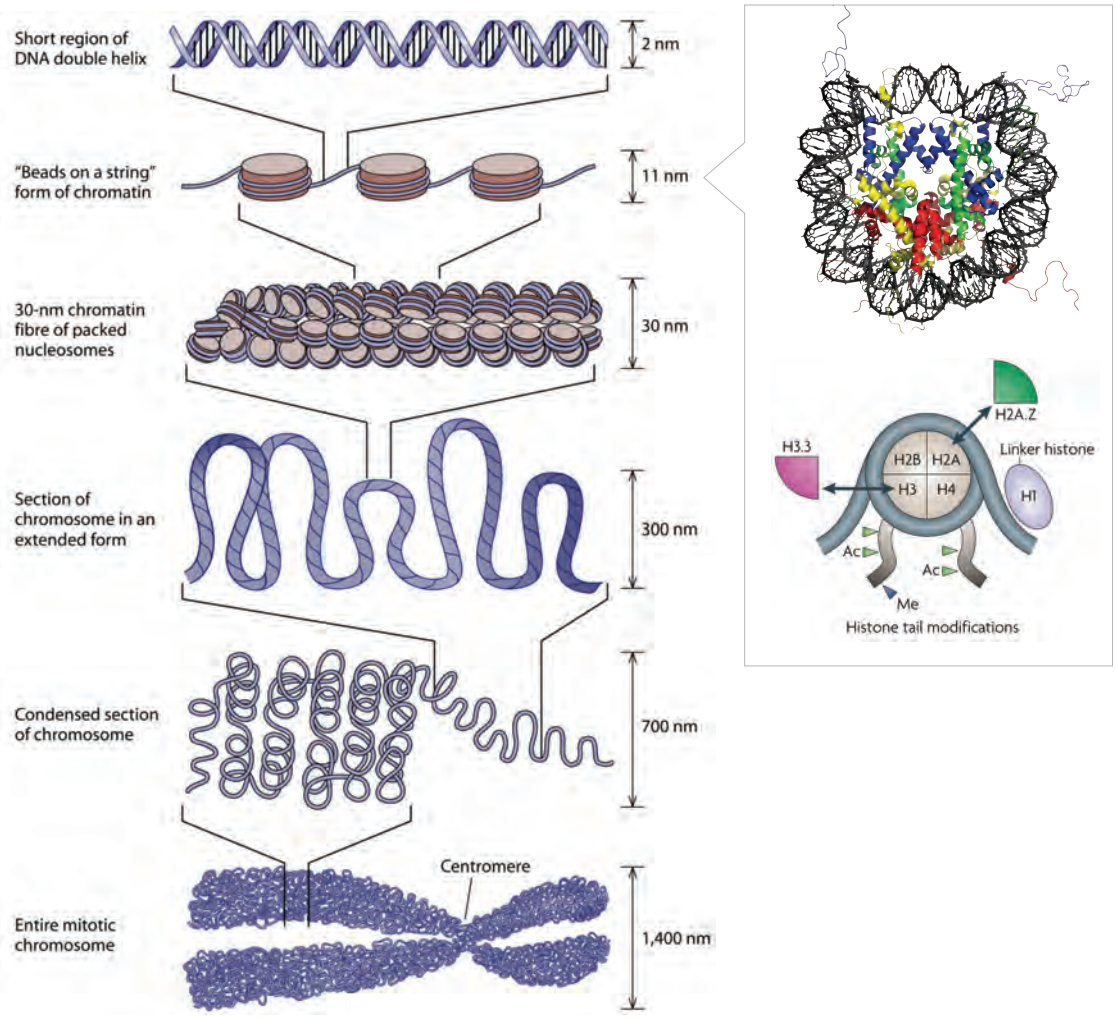


Figure 1.1. Packages of DNA.

(*Left panel*) Hierarchical organization of eukaryotic chromatin. DNA is wrapped around a histone octamer to form nucleosomes. Nucleosome array spaced by linker DNA can be folded into a fiber-like structure of ~30 nm in diameter. Chromatin is then compacted into high-order structures such as euchromatin and heterochromatin in interphase or mitotic chromosome during mitosis. (*Right panel*) Crystal structure of the nucleosome core particle. About 147 bp of DNA wrap around a histone octamer shown as two copies of H2A (yellow), H2B (red), H3 (blue), and H4 (green). The H3/H4 dimers are joined to form a tetramer and mark the center of the nucleosome or dyad of the nucleosome. The H2A/H2B dimers attach via interaction between H2A and H4 to form the octamer.

Left panel: Adapted from (FELSENFELD AND GROUDINE 2003) & (JANSEN AND VERSTREPEN 2011). Cartoon shows the histone variants H2A.Z and H3.3, histone H1 bound to linker DNA, and covalent modifications on histone tails.

- Left panel: Reprinted and adapted with permission from American Society for Microbiology: (JANSEN AND VERSTREPEN 2011).
- Right panel: Reprinted and adapted with permission from Nature Publishing Group: (LUGER et al. 1997) for the crystal structure of nucleosome & (JIANG AND PUGH 2009) for the cartoon of nucleosome.

Primary chromatin structures at a glance

Nucleosomes: position, position, position

The nucleosome is the basic, repeating unit which forms chromatin in eukaryotic cells. It consists of ~147 bp of DNA wrapped around histone octamers built from pairs of histones H2A, H2B, H3, and H4 (KORNBERG 1974; LUGER *et al.* 1997; KORNBERG AND LORCH 1999). A series of nucleosomes packaging a long stretch of genomic DNA was readily observed as beads-on-a-string structure by EM studies (OLINS AND OLINS 1974). At this scale, a great deal of interest surrounds the biological meaning of nucleosome occupancy along the genome. Where are nucleosomes well-positioned or depleted? Which nucleosomal regions are fragile or highly dynamic? What is the functional consequence of nucleosome occupancy with a certain pattern? The principle methodology used to approach these questions is based on the differential susceptibility of DNA to various fractionation processes, with nucleosomal DNA being protected from enzymatic digestion or chemical attack, and linker DNA being relatively accessible (**Figure 1.2**) (ZENTNER AND HENIKOFF 2014). Many genome-wide assays have been

developed for mapping different features of primary chromatin structure. For example, nucleosome depletion regions (NDRs) (typically found near regulatory regions such as promoters and enhancers) can be profiled by a category of enzyme-based methods such as DNase-seq (DNase I) (KEENE *et al.* 1981; MCGHEE *et al.* 1981; CRAWFORD *et al.* 2006a; CRAWFORD *et al.* 2006b; BOYLE *et al.* 2008), ATAC-seq (Tn5 transposase) (BUENROSTRO *et al.* 2013), or NOMe-seq (GpC DNA methylase) (KELLY *et al.* 2012), which preferentially target areas with high accessibility. Additionally, given the different levels of nucleosome and protein occupancy on the genome, methods such as FAIRE-seq (Formaldehyde-Assisted Isolation of Regulatory Elements) are capable of physically separating NDRs from bulk chromatin, as non-crosslinked DNA largely is fractionated into the aqueous phase, while protein-crosslinked fragments are retained in the organic phase (NAGY *et al.* 2003).

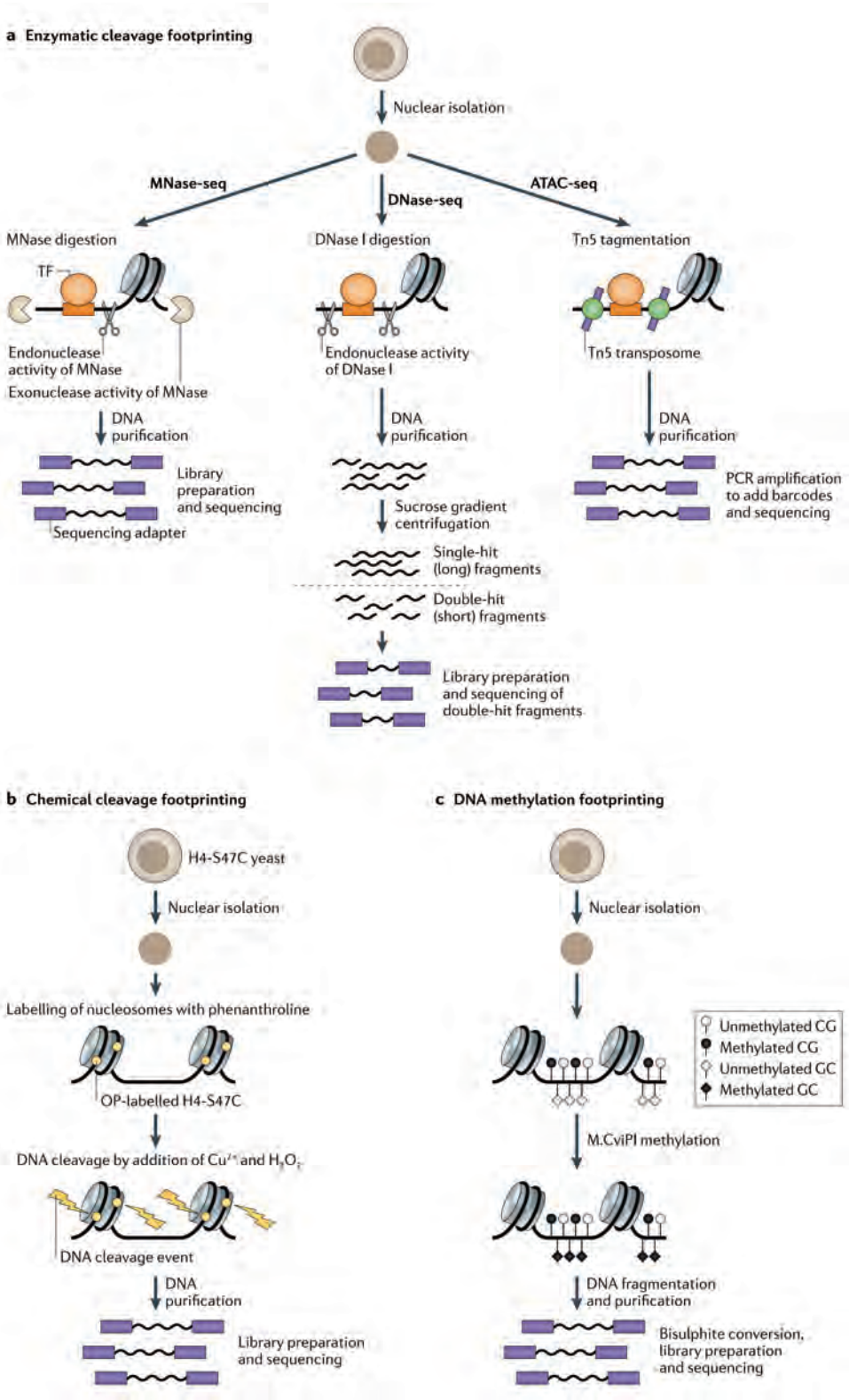

a Enzymatic cleavage footprinting
Nuclear isolation
MNase-seq
DNase-seq
ATAC-seq
MNase digestion
DNase I digestion
Tn5 tagmentation
TF
Endonuclease activity of MNase
Exonuclease activity of MNase
Endonuclease activity of DNase I
Tn5 transposome
DNA purification
DNA purification
DNA purification
Library preparation and sequencing
Sequencing adapter
Sucrose gradient centrifugation
Single-hit (long) fragments
Double-hit (short) fragments
Library preparation and sequencing of double-hit fragments
PCR amplification to add barcodes and sequencing
b Chemical cleavage footprinting
H4-S47C yeast
Nuclear isolation
Labelling of nucleosomes with phenanthroline
OP-labelled H4-S47C
DNA cleavage by addition of Cu2+ and H2O2
DNA cleavage event
DNA purification
Library preparation and sequencing
c DNA methylation footprinting
Nuclear isolation
Unmethylated CG
Methylated CG
Unmethylated GC
Methylated GC
M.CviPI methylation
DNA fragmentation and purification
Bisulphite conversion, library preparation and sequencing

Figure 1.2. Methods for chromatin footprinting.

(a) For enzymatic cleavage methods, nuclei are isolated and treated with the enzyme of interest to fragment chromatin, and DNA is purified for sequencing. (b) For chemical cleavage method, isolated nuclei harboring a single copy of H4S47C mutation are labeled with (N-(1,10-phenanthroline-5-yl)iodoacetamide) (OP) and treated with Cu2+ ions and hydrogen peroxide (H2O2) to cleave DNA. DNA is then isolated for sequencing. (c) For DNA methylation footprinting, isolated nuclei are treated with M.CviPI to methylate cytosine in the GC dinucleotide context. DNA is fragmented, purified, and subjected to bisulphite conversion. Resulting DNA is then sequenced.

- Reprinted and adapted with permission from Nature Publishing Group: (ZENTNER AND HENIKOFF 2014).

Our current concept of nucleosome occupancy and positioning are illustrated mainly via the wide-ranging application of micrococcal nuclease (MNase) digestion on chromatin, which has higher preference for DNA linkers than nucleosomal DNA (NOLL AND KORNBERG 1977). A desired level of chromatin fragmentation can be tightly controlled by a series of MNase titrations, usually resulting in a regular pattern of DNA ladder with repeating nucleosomal sizes (e.g., mononucleosomes, dinucleosomes, and so on). Sequencing DNA at mononucleosome size reveals genomic positions of three types of nucleosome: 1) well-positioned nucleosomes; 2) fuzzy nucleosomes; and 3) fragile nucleosomes (**Figure 1.3**) (HUGHES AND RANDO 2014). The well-positioned nucleosomes that reside in the same location in most of cell populations are highly enriched at the borders of regulatory elements. Nucleosome positioning becomes fuzzier moving away from well-positioned nucleosomes (> 1500 bp), reflecting varying locations between cells. A low-level of MNase digestion further

reveals a subset of fragile or high turnover nucleosomes located in regulatory regions (WEINER *et al.* 2010; XI *et al.* 2011), which is consistent with recent result of chemical cleavage studies (BROGAARD *et al.* 2012; VOONG *et al.* 2016). In addition, the locations of DNA binding proteins such as transcription factors can be mapped by collecting small sized DNA fragments (typically 35 – 100 bp) after 'light' MNase digestion (HENIKOFF *et al.* 2011; KENT *et al.* 2011).

Genetic and biochemical studies provide insights into the forces driving the organization of nucleosome positions (**Figure 1.3**) (KORBER AND HORZ 2004; SEKINGER *et al.* 2005; KAPLAN *et al.* 2009; ZHANG *et al.* 2011; VAN BAKEL *et al.* 2013). First, the flexibility of the DNA that wraps histones is determined by its base composition. Highly stiff poly(dA/dT) tracts are disfavored for nucleosome incorporation and thus are enriched in nucleosome depletion regions. Second, steric inhibition with adjacent nucleosomes or DNA binding proteins also influence nucleosome positions, as predicted by a statistical positioning model (KORNBERG AND STRYER 1988). Third, the ATP-dependent chromatin regulators such as RSC, INO80, and ISWs can remodel the primary chromatin landscape by assembling, evicting, and moving nucleosomes (CLAPIER AND CAIRNS 2009). Thorough dissection of these mechanics was done in a recent study using a biochemical *in vitro* system (KRIETENSTEIN *et al.* 2016). Fourth, Pol2 readthrough can destabilize nucleosome positioning, as evidenced by highly transcribed genes having lower and fuzzier nucleosome occupancy (TSANKOV *et al.* 2010). Finally, structural proteins, such as Heterochromatin Protein 1 (HP1) and histone

H1, can overcome steric hindrance and assemble more constrained structures. In sum, the combination of these forces delineates the locations of NDRs and boundaries, and guides wrapping of the remaining region, which leads to the first dimensional repeating structure.

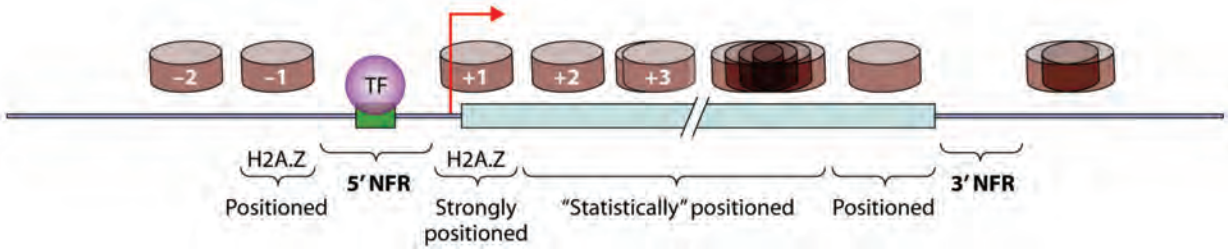


Figure 1.3. Chromatin landscape.

Nucleosome depletion regions (NDRs) are mainly defined by stiff poly(A/T) tracts, in which the incorporation of nucleosome is extremely unstable. These regions are usually found near regulatory elements (promoters and enhancers) and bound by transcription factors or fragile nucleosomes. NDRs are typically surrounded by well-positioned nucleosomes, which are localized at -1 and +1 positions of genes and enriched with H2A.Z variant. The formation of nucleosomes is further enhanced by DNA sequences containing regularly spaced A/T dinucleotides (approximately one dinucleotide every 10 bp) and G/C dinucleotides (in between the A/T dinucleotides). Nucleosome are also positioned by the steric hindrance of neighboring nucleosomes. This imperfect positioning signal results in the precision of nucleosome positioning being gradually decayed, and nucleosome positioning becomes fuzzier around 1.5kb downstream of the +1 nucleosomes. Other factors like ATP-dependent chromatin remodelers can evict, slide, or assemble nucleosomes, adding another layer of regulation of chromatin landscape.

- Reprinted and adapted with permission from American Society for Microbiology: (JANSEN AND VERSTREPEN 2011).

Histone variants and modifications

The repeating nucleosomal structure is not composed of uniform 'beads.' Individual nucleosomes can be decorated with various covalent modifications and subunits, and the chemical properties of 'beads' are differed greatly from each other. For example, histone variants (e.g., H2A.X and H2A.Z isoforms of H2A and H3.1 and H3.3 isoforms of H3) can undergo dynamic exchange with canonical histones on the repeating nucleosome template (**Figure 1.1**). The tails and core domain of histones can be chemically modified with a great variety of modifications such as acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, and more at multiple residues (**Figure 1.4**) (TAN *et al.* 2011). These modifications and variants alter the physical and chemical properties of nucleosomes and their microenvironment, resulting in the modulation of enzyme binding affinity, and thus intimately affecting the consequences of genomic transactions (TURNER 1993; JENUWEIN AND ALLIS 2001; RANDO 2012).

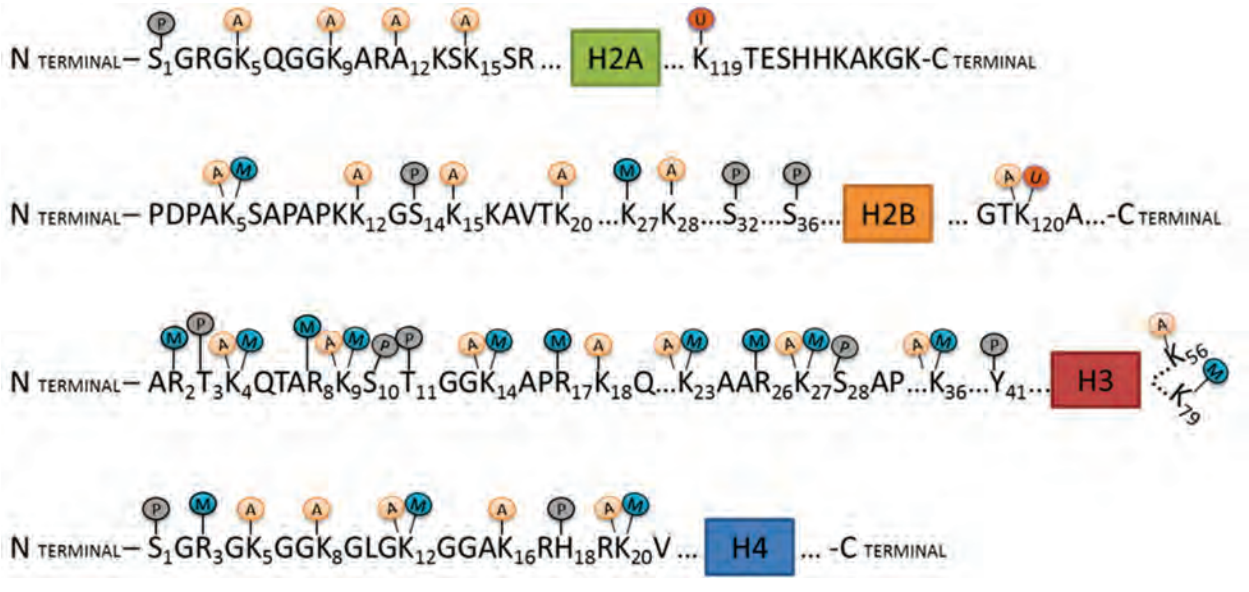


Figure 1.4. Overview of histone modifications.

Schematic shown the common chemical modifications on histone tails and cores, including acetylation, methylation, phosphorylation, ubiquitination, sumoylation, etc.

- Reprinted and adapted with permission from InTech: (XU et al. 2013).

Genome-wide mapping of histone modifications largely relies on the methods of chromatin immunoprecipitation (ChIP-seq or ChIP-Chip) following chromatin fragmentation by MNase digestion or mechanical sonication. Over a decade of efforts revealed the key aspects of histone modifications conserved across species (**Figure 1.5**) (LIU *et al.* 2005; SINHA *et al.* 2006; FILION *et al.* 2010; ERNST *et al.* 2011). First, the process of transcription including Pol2 initiation and elongation leaves a massive footprint on chromatin. This results in the deposition of di/trimethylation on H3K4 (H3K4me2/3), multiple acetyl groups on histone tails (H3K4/9/14/18/27ac and H4K5/8/12ac), and H3.3 and H2A.Z isoforms at the 5'-ends of genes, while histone tails at the 5'-end of poised genes are marked with relatively lower levels of acetylation. Additionally, nucleosomes across the gene body are broadly marked with H3K36me3 and H3K79me3, in company with decreasing levels of acetylation (RANDO AND WINSTON 2012). Second, distal regulatory elements such as enhancers are marked with H3K4me1/2, along with modifications at H3K27 (acetylation or methylation) which can be used to further distinguish repressed, poised, or active enhancers (HEINTZMAN *et al.* 2007). Third, constitutive heterochromatin (e.g., telomeric and repetitive regions) are enriched with H3K9me3. In contrast, facultative heterochromatin in which the genes are repressed by polycomb complexes are marked with H3K27me3. A

subset of areas commonly found at developmental loci such as Hox clusters in embryonic stem cells are simultaneously marked as the bivalent state, with the active mark H3K4me1 and the repressive mark H3K27me3, to allow for a quick response to developmental signals (BERNSTEIN *et al.* 2006). Finally, the centromeric regions are usually enriched with the H3-like CENP-A protein and broadly marked with H3S10ph during mitosis. Overall, the states of nucleosomes strikingly mirror genomic functions, which has been broadly used as the standard reference to predict and model chromatin states.

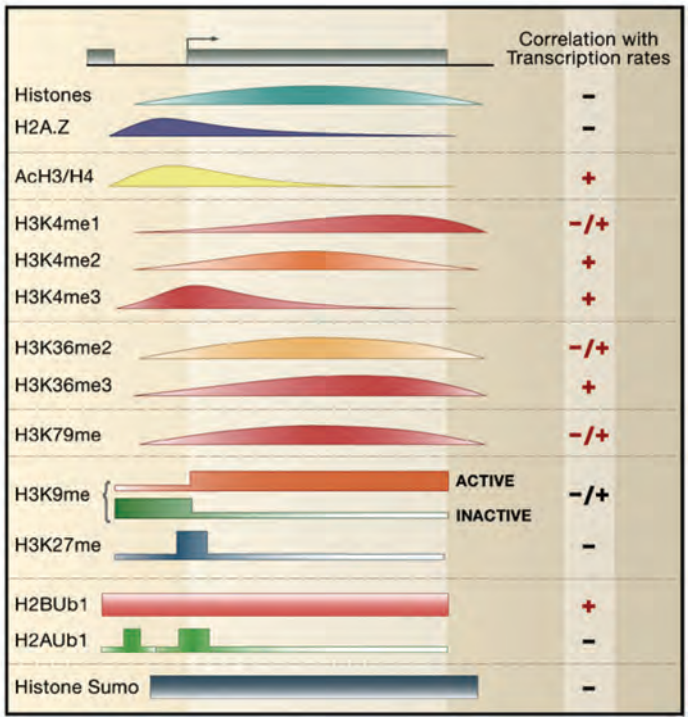


Figure 1.5. Genome-wide distribution pattern of histone modifications from a transcription perspective.

Schematic shown the genome-wide average of histone modifications across an arbitrary area containing 5'- and 3'-ends of nucleosome free regions (NFRs), transcription start

site (TSS), and gene body. The curves are showing the genome-wide average patterns of histone modifications from yeast, except for H3K9me3 and H3K27me3.

- Reprinted and adapted with permission from Elsevier: (LI et al. 2007).

Although histone modifications have been investigated intensively in the past, the kinetics of their modifications, the cause-or-consequence relationships of many genomic functions, and even the plausibility of a meaningful 'histone code' remain under debate. A recent modified version of ChIP (called co-ChIP) which includes serial pulldowns of target modifications seems promising in an effort to directly dissect the co-existence of chemical modifications on individual nucleosomes in more detail (SADEH *et al.* 2016; WEINER *et al.* 2016). Another innovative approach called chromatin endogenous cleavage (ChEC-seq) appears to be able to efficiently reduce background noise compared to traditional ChIP, and whose further application in the mapping of histone modifications will provide a complementary view of chromatin states (ZENTNER *et al.* 2015).

Controversial 30 nm chromatin structures

In 1976, Aaron Klug and colleagues folded purified nucleosomal arrays *in vitro* into regular 30 nm chromatin fibers (FINCH AND KLUG 1976), pioneering the entire era of 30 nm chromatin hunting. Despite over 40 years of research, how nucleosomes interact with each other and whether these interactions configure any periodic chromatin structure remains elusive (TREMETHICK 2007; MAESHIMA *et al.* 2010a; FUSSNER *et al.* 2011; LUGER *et al.* 2012). Decades of biochemical

studies combined with electron microscopy and other biophysical methods (e.g., X-ray diffraction, optical/magnetic tweezer, mesoscopic modeling, etc.) have proposed two competing models for 30 nm chromatin folding (**Figure 1.6A-B**): 1) the one-start helix or solenoid path of folding, and 2) the two-start helix or zigzag path of folding (ROBINSON AND RHODES 2006; LI AND REINBERG 2011; GRIGORYEV AND WOODCOCK 2012; LUGER *et al.* 2012). The 'one-start' solenoid is arranged by consecutive nucleosomes with approximately six nucleosomes per turn, in which nucleosomes interact with each other and follow a helical trajectory with bending of linker DNA (FINCH AND KLUG 1976). In the 'two-start' zigzag model, two ribbons of nucleosomes form a two-start helix linked by relatively straight linker DNA so that alternate nucleosomes become adjacent partners with roughly three nucleosomes per turn (WORCEL *et al.* 1981; WOODCOCK *et al.* 1984).

30 nm chromatin in vitro

The main impediment to solving the 30 nm chromatin structure is the high-degree of complexity in native chromatin, which contains variable lengths of linker DNA, intricate combinations of histone modifications and variants, chromatin remodelers, and transcription factors, all of which greatly complicate downstream analysis and interpretation. The extremely compact nature of the structure in nucleus further prevents the regular path of chromatin from being visualized by any given method. To simplify experimental variables, the majority of studies used an artificial chromatin template, typically 12 identical repeats of 'Widom-

601' nucleosome positioning sequence that can be assembled into nucleosome arrays with highly uniform recombinant histones (LOWARY AND WIDOM 1998). For example, Richmond and colleagues designed two prominent studies, one which crosslinked nucleosome interactions via modified histone H4 and H2A (DORIGO *et al.* 2004), and the other crystallized the tetranucleosome array (**Figure 1.6C**) (SCHALCH *et al.* 2005). Both strongly suggested the zigzag path of chromatin folding. In contrast, later cryo-EM studies on long nucleosomal arrays with various lengths of nucleosomal repeats and in the presence of histone H1, preferred the multi-start interdigitated solenoid model, as the high nucleosome packing ratio was not compatible with any two-start zigzag or one-start solenoid model (ROBINSON *et al.* 2006; WONG *et al.* 2007; ROUTH *et al.* 2008). Single-molecule force spectroscopy (SMSF) also can be used to model the internal structure of 30 nm chromatin fiber by applying a pulling force on bead-conjugated chromatin fibers via optical/magnetic tweezers. The resulting kinetics on fiber extension can be further interrelated to different chromatin structures – two independent studies argued for chromatin folding consistent with the zigzag model at physiological concentration of monovalent cations (CUI AND BUSTAMANTE 2000) but with the solenoid model in the presence of divalent cations (KRUITHOF *et al.* 2009), rather than the multi-start interdigitated structure observed by cryo-EM.

More recently, Schlick and colleagues used electron microscopy-assisted nucleosome capture (EMANIC) combined with mesoscopic modeling, which provided an alternative perspective for the 30 nm chromatin structures, at least for *in vitro*-assembled fibers (GRIGORYEV *et al.* 2009). In the method, condensed chromatin was first crosslinked by formaldehyde, decondensed in low salt conditions, and then subjected to transmission EM imaging to define the patterns of inter-nucleosome interactions. Surprisingly, instead of a uniform structure, the observation of heterogeneous types of nucleosome interactions strongly suggested heteromorphic folding for 30 nm chromatin, with a predominant two-start zigzag path interspersed with one-start solenoid structure with partially bent linker DNA (**Figure 1.6D**) (GRIGORYEV *et al.* 2009). *In silico* simulation such as mesoscopic modeling also supports a heteromorphic conformation, as it is energetically more favorable than the uniform structures under high levels of compaction (ARYA AND SCHLICK 2006; PERISIC *et al.* 2010; COLLEPARDO-GUEVARA AND SCHLICK 2012). Longer nucleosome repeat lengths (> 200 bp) often require linker histone H1 to prevent DNA bending, thus disfavoring the formation of a solenoid helix. The features of a heteromorphic model fully agree with the nature of chromatin fibers observed by EM, as well with the length in diameter and mass-per-unit of chromatin isolated from nuclei.

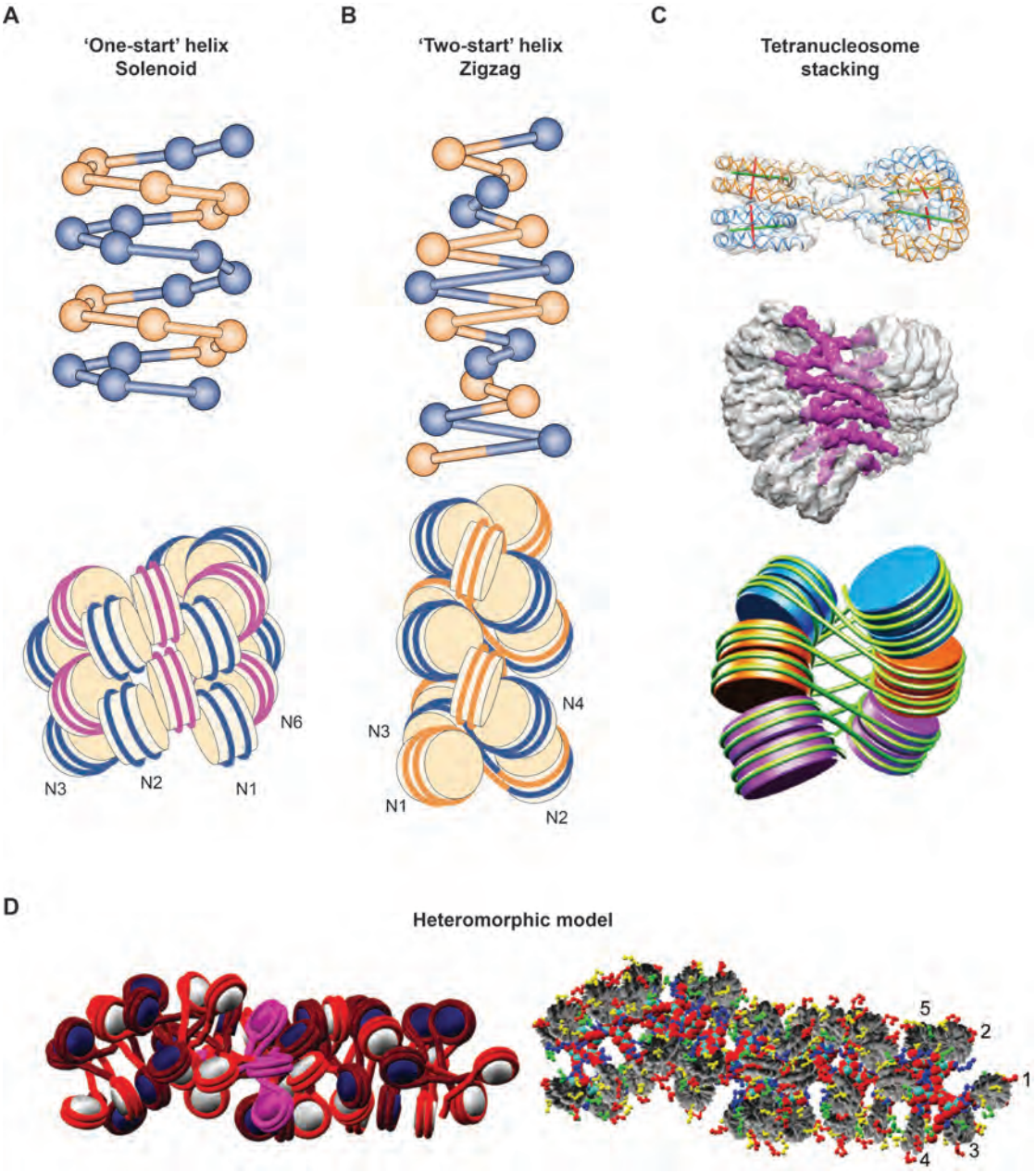


Figure 1.6. Models for chromatin secondary structures.

(**A**) The solenoid model is characterized by a consecutive nucleosome helix in which a nucleosome in the fiber tends to repeatedly interact with its fifth and sixth neighbor nucleosomes (N+4 and N+5). Solenoid fibers consist of longer NRL ~197bp, with 5 – 6 nucleosomes per helical turn and ~30 nm of diameter. (**B**) In the zigzag structure, a nucleosome prefers to interact with its second neighbor nucleosome (N+2), thus forming

the zigzag path of chromatin with roughly 3 nucleosomes per turn. The NRL and the diameter of zigzag structure are shorter than the solenoid structure, with ~167bp per repeating unit and ~25nm of diameter. (**C**) The tetranucleosome stacking model from cryo-EM data. The top panel shown the segmented density map for the tetranucleosome unit. The structure is highly similar to the crystal structure of tetranucleosome motif with 167bp of NRL in the absence of histone H1. The results indicated that the presence of histone H1 and variable lengths of NRL do not alter interactions within tetranucleosome, but affects the separation and rotation between two stacks. Axes are highlighted by green and red lines. The middle panel shown the stacking of three tetranucleosome units forming the two-start 30 nm chromatin fiber. The bottom panel shown the schematic of the middle panel. (**D**) The heteromorphic chromatin model. (*Left panel*) The results of EMANIC data and mesoscopic modeling proposed the heteromorphic model for chromatin secondary folding, with a predominate zigzag path interspersed with solenoid path. (*Right panel*) A detailed view of heteromorphic model shown in the left panel. The histone core particles are rendered with gray color, and the tails are colored yellow for H2A, red for H2B, blue for H3, and green for H4. The first five nucleosomes are indicated.

- Panel A-B: Reprinted and adapted with permission from Nature Publishing Group: (LUGER et al. 2012).
- Panel C: Reprinted and adapted with permission from AAAS: (SONG et al. 2014).
- Panel D: Reprinted and adapted with permission from ASBMB: (SCHLICK et al. 2012).

In addition to the *in vitro* synthesis system, chromatin isolated from nuclei and followed by nuclease fragmentation (usually in the range of one to several hundred nucleosomes) has been used to study 30 nm chromatin structures as well (THOMA *et al.* 1979). Isolated fibers partially retain the complex nature of native chromatin that mimics various nucleosome repeat lengths and histone compositions/modifications *in vivo*. These '*in vivo*-like' fibers have been most effective in controlling salt-dependent chromatin folding, as mentioned in the first

30 nm chromatin model (FINCH AND KLUG 1976). Nevertheless, the high complexity of *in vivo*-like chromatin fibers often obscured downstream interpretation of results from EM or atomic force microscopy (WOODCOCK *et al.* 1993; LEUBA *et al.* 1994; BEDNAR *et al.* 1998), which suggests an intrinsic heterogeneity of nucleosome packing within 30 nm chromatin fibers resulting from the variability in nucleosome orientation. Interestingly, a recent EMANIC study on isolated interphase chromatin using HeLa cells and chicken erythrocytes supported the heteromorphic chromatin structure with a predominant zigzag path (GRIGORYEV *et al.* 2016).

In brief, a multitude of key factors have been reported to influence folding of the linear 'beads-on-a-string' polymer into compact secondary structures, including: 1) the concentration of salt (NaCl), divalent ions (Mg2+, Ca2+), and polyamines (spermidine, spermine); 2) the N-terminal histone tails and their chemical modifications (H4 tail and acidic patches); 3) nucleosome repeat length (NRL); 4) the level of histone H1 binding on chromatin; 5) chromatin regulators (e.g. Polycomb or Sir complex); and 6) molecular crowding effects (dextran, PEG).

30 nm chromatin in vivo

Observation of chromatin structure *in situ* has been an immensely challenging task due to the high-degree of compaction in the nucleus. Only when nuclei were mildly decondensed or spread on a water surface can EM imaging distinguish

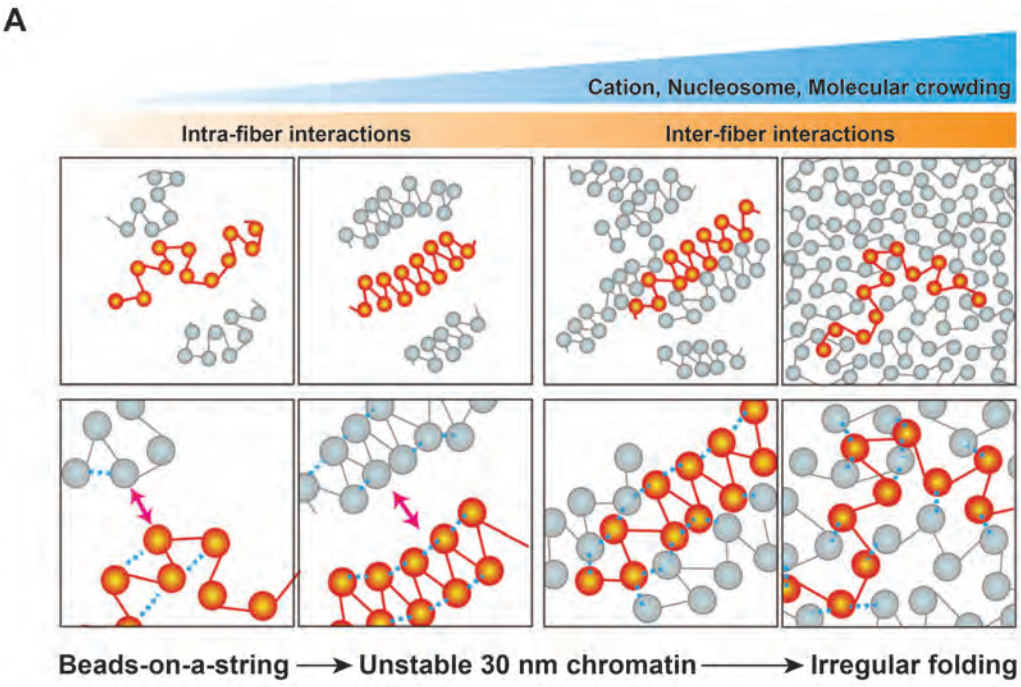
chromatin structures from background signals of dense mass (GALL 1963; GALL 1966). *In situ* 30 nm chromatin was found exclusively in some 'special' types of nuclei, such as nucleated erythrocytes in chicken or necturus (WILLIAMS *et al.* 1986; SCHEFFER *et al.* 2011; GRIGORYEV *et al.* 2016), sperm in Echinoderm (starfish), Thyone briareus (sea cucumber), and Patiria miniata (bat star) (WILLIAMS *et al.* 1986; HOROWITZ *et al.* 1994; SCHEFFER *et al.* 2012), and Xenopus laevis eggs (GALL 1963; GALL 1966), whose structure mirrored the zigzag conformation identified by *in vitro* methods. Another special case in mouse retinal rod photoreceptors revealed 30 nm chromatin fibers at the outer layer of centric heterochromatin, where chromatin is less compact (KIZILYAPRAK *et al.* 2010). It is worth noting that the shared features in these types of nuclei may explain the discovery of 30 nm chromatin, including 1) the absence of active transcription, 2) a specialized type of histone H1, 3) very long nucleosome repeat lengths, and 4) a low proportion of non-histone chromatin proteins.

Besides microscopic approaches, a theoretical method of modeling 30 nm chromatin structures was pioneered by Chatterjee and colleagues, who induced spatially correlated DNA breaks by passing gamma rays through aqueous samples (RYDBERG *et al.* 1998). The pattern of averaged fragment length distribution (FLD) resulting from the periodic folding of DNA strongly argued that the predominant two-start zigzag path of chromatin folding are widespread structures in mammalian cells. This experimental concept recently was combined with high-throughput sequencing, which further highlighted that zigzag folding

was enriched in H3K9me3 regions, while H3K27me3 and open chromatin contained a mixture of solenoid and linear conformations (RISCA *et al.* 2017).

However, a growing number of research provides evidence that chromatin consists of irregular 10 nm fibers *in vivo*, rather than a periodic 30 nm structure (TREMETHICK 2007; MAESHIMA *et al.* 2010a; MAESHIMA *et al.* 2010b; FUSSNER *et al.* 2011; JOTI *et al.* 2012; LUGER *et al.* 2012; MAESHIMA *et al.* 2016). Cryo-EM and electron spectroscopic imaging (ESI) studies in mammalian interphase or metaphase nuclei suggested bulk chromatin is folded into disordered 10 nm fibers and non-uniform sizes of nucleosomal clutch but not in any type of periodic structure (ELTSOV *et al.* 2008; AHMED *et al.* 2010; FUSSNER *et al.* 2012; CHEN *et al.* 2016). Regular chromatin structures only appeared under conditions where nuclei were artificially pre-decondensed in low salt solutions or crosslinked by aldehyde reagents (ELTSOV *et al.* 2008). Moreover, small-angle X-ray scattering (SAXS) analysis only detected peaks at ~6 nm (face-to-face stacking) and ~ 11 nm (edge-to-edge stacking) showing predominance of 10 nm chromatin fibers in mammals, while the ~30 nm peak readily disappeared as ribosomes were washed away (NISHINO *et al.* 2012). These results again strongly argued against the existence of 30 nm chromatin *in vivo*. Recent super-resolution imaging (STORM) revealed heterogeneous groups of nucleosomal clutches ubiquitously exhibited in nuclei, and whose sizes strongly correlated to stem cell states and local transcriptional activity (RICCI *et al.* 2015). Interestingly, although no periodic chromatin folding was clearly identified, the structure of nucleosome clutches

appears to mirror the tri- or tetra- nucleosome motif observed by X-ray crystallization or cryo-EM (SCHALCH *et al.* 2005; SONG *et al.* 2014), and likely act as the building block of chromatin. In sum, the current model suggests liquid-like behavior of chromatin folding (**Figure 1.7**) (LUGER *et al.* 2012; MAESHIMA *et al.* 2016). The 30 nm fiber may only occur *in vitro* due to the high dilution of chromatin fibers used in such studies, in which a given nucleosome will only have access to other nucleosomes on the same DNA fragment (intra-fiber interactions > inter-fiber interactions). While in the nuclear "sea of nucleosomes" many additional nucleosomes are available in *trans* for internucleosomal interactions fragment (inter-fiber interactions > intra-fiber interactions), the irregular globule structure ultimately predominates periodic chromatin folding.


A
Cation, Nucleosome, Molecular crowding
Intra-fiber interactions
Inter-fiber interactions
Beads-on-a-string
Unstable 30 nm chromatin
Irregular folding

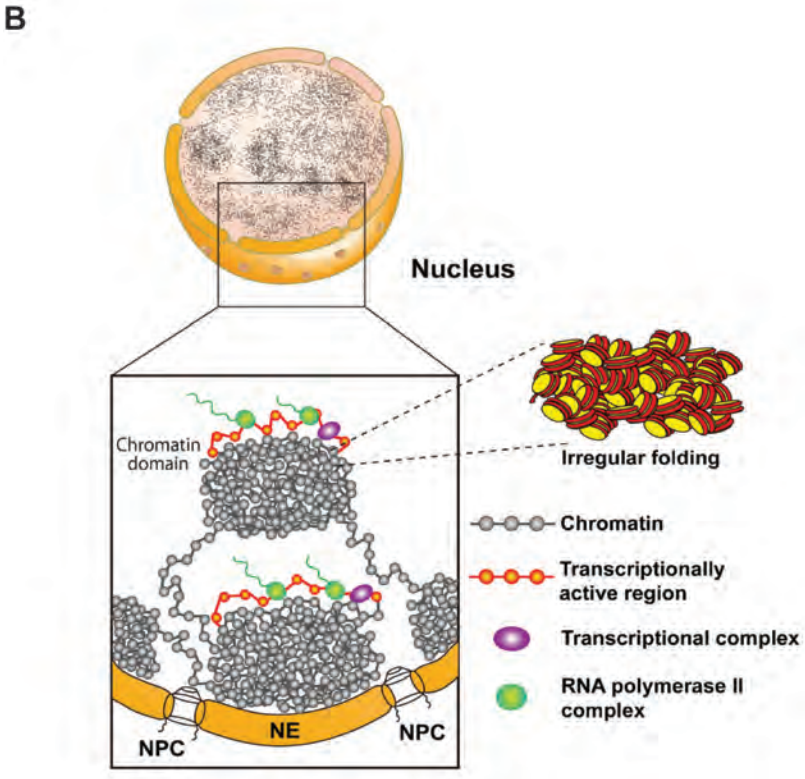

B
Nucleus
Chromatin domain
Irregular folding
Chromatin
Transcriptionally active region
Transcriptional complex
RNA polymerase II complex
NPC
NE
NPC

Figure 1.7. The liquid-like chromatin model.

(A) In the dilution condition, the predominant intra-fiber force may facilitate the fiber folding into the 30 nm chromatin structures. However, the irregular chromatin fibers readily become the dominant structures with increasing strength of inter-fiber interactions being the major force. The chromatin structures are sensitive to the change of the concentration of cation, histone H1, and molecular crowding, as well as the state of nucleosome modifications. (B) Schematic shown repressed chromatins are organized as viscous liquid-like drops, which formation is mediated by the inter-nucleosome interactions and macromolecular crowding effects. These viscous drops are linked (or delimited) by flexible chromatin fibers. Highly-transcribed regions are looped out from the edge of repressed regions, in which transcription factors and Pol2 are accessible to the chromatin context. The concept of the liquid-like chromatin model fully agrees with many features of the current chromatin conformation model proposed via Hi-C (see next section).

- Reprinted and adapted with permission from Springer: (MAESHIMA et al. 2014).

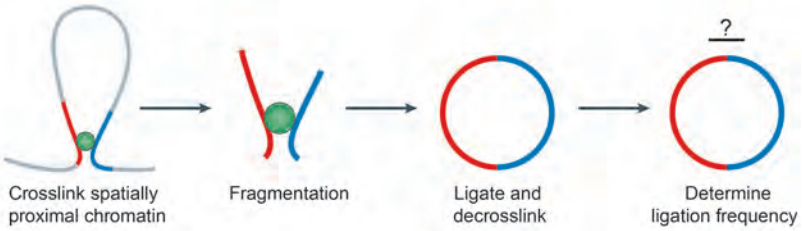
Beyond the 30 nm chromatin fiber

Chromosome Conformation Capture

Structural analysis of chromosome folding beyond the nucleosome fiber has been revolutionized by the Chromosome Conformation Capture (3C) family of techniques, which measure relative contact frequency between pairs of genomic loci *in vivo* (DEKKER *et al.* 2002; GIBCUS AND DEKKER 2013; BONEV AND CAVALLI 2016; SCHMITT *et al.* 2016). In 3C-based protocols, chromatin is first crosslinked *in vivo* using formaldehyde to capture physical interactions between distal regions of the genome. Chromatin subsequently is fragmented by restriction enzymes, and ligation of chromatin fragments is used to generate chimeric DNA molecules (DEKKER *et al.* 2002). Detecting these molecular libraries provides a

readout of genomic loci that were crosslinked to one another via protein-protein interactions (**Figure 1.8A**). An increasing number of 3C variants have been developed based on needs of widespread experimental designs, with scales of data throughput that include methods for detecting single-to-single (3C) (DEKKER *et al.* 2002), single-to-all (4C) (SIMONIS *et al.* 2006; ZHAO *et al.* 2006), multiple-to-multiple (5C) (DOSTIE *et al.* 2006), and all-to-all (Hi-C) (LIEBERMAN-AIDEN *et al.* 2009; DUAN *et al.* 2010; KALHOR *et al.* 2011), in some cases can be applied for single-cell study (NAGANO *et al.* 2013; RAMANI *et al.* 2017). In addition, methods that specifically enrich interactions by targeting proteins of interest (ChIA-PET) or genomic loci of interest (Capture-C) can be used broadly for studying protein-mediated chromatin structures or promoter-promoter/promoter-enhancer interactions (**Figure 1.8B**) (FULLWOOD *et al.* 2009; HUGHES *et al.* 2014). Genome-wide variants of 3C, such as Hi-C, have revealed a number of organizational features of the eukaryotic genome at increasingly finer resolutions (**Figure 1.8C**), from the scale of full chromosomal territories, to multi-megabase active and inactive compartments (LIEBERMAN-AIDEN *et al.* 2009), to hundred-kilobase topologically-associating domains (TADs) (DIXON *et al.* 2012; NORA *et al.* 2012), and to long-range chromatin loops (PHILLIPS-CREMINS *et al.* 2013; SOFUEVA *et al.* 2013; RAO *et al.* 2014), whose regulatory mechanisms and biological consequences have been subjected to intensely investigate in the recent years.

A



B

Technique	Purpose	Principle of detection
3C	One-to-one	qPCR
4C	One-to-all	Inverse PCR from selected fragment plus sequencing
Capture -C	One/many-to-all	Pull down of selected fragment plus sequencing
5C	Many-to-many	Multiplex ligation-mediated amplification plus sequencing
ChIA-PET	Many-to-many	ChIP plus sequencing
Hi-C	All-to-all	Sequencing

C

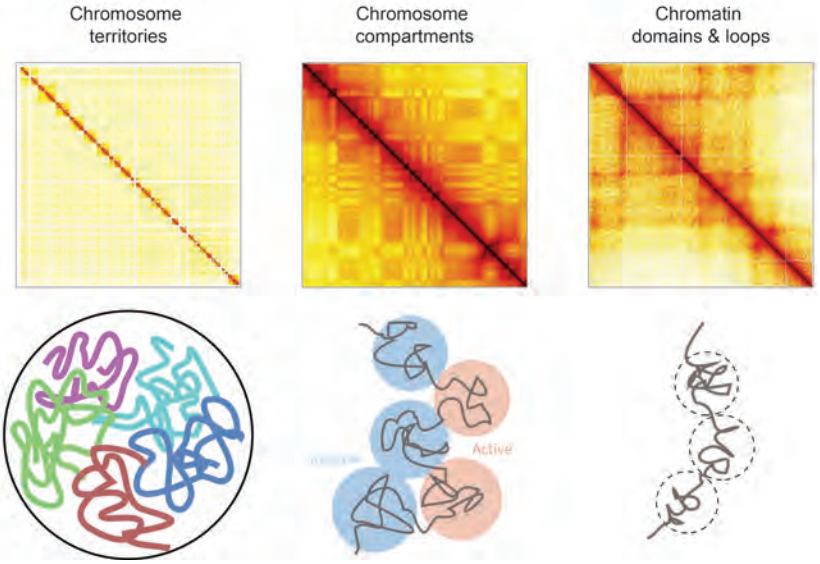


Figure 1.8. Overview of 3C-based methods.

(**A**) Schematic of a simplified procedure for 3C-based methods. The key steps include crosslinking chromatin, genome fragmentation, proximity ligation, and detection of the ligated products. (**B**) Table of the current 3C-family techniques. Many 3C-based methods have been developed based on the different detection methods, ranging from the scale of a single locus to genome-wide. (**C**) An example of Hi-C data. The contact matrices and cartoons show the identified chromosome structures such as whole genome territories, chromosome compartments, TADs, and chromatin loops with increasing resolution of view.

- Panel A-B: Reprinted and adapted with permission from Nature Publishing Group: (KRIJGER AND DE LAAT 2016).
- Panel C: Reprinted and adapted with permission from Geoffrey Fudenberg, (2017): 03-23-17_les_houches_forWeb_gfudenberg.pdf. figshare. https://doi.org/10.6084/m9.figshare.4871948.v1.

Chromosome compartments

Eukaryotic cells non-randomly compartmentalize the genome into euchromatin and heterochromatin, which has been extensively observed by using microscopic approaches such as DNA fluorescence *in situ* hybridization (FISH) that cytologically display the co-distribution of regions with similar genomic features (SCHERMELLEH *et al.* 2001; ALBIEZ *et al.* 2006; SHOPLAND *et al.* 2006; HU *et al.* 2009; ROUQUETTE *et al.* 2009; CREMER AND CREMER 2010; HUBNER *et al.* 2015). In other words, gene dense regions are preferentially co-localized in area with active transcription and, in general, gene deserts are contained within compact chromatin or at the nuclear periphery. Differential enrichments of gene density between euchromatin and heterochromatin broadly has been observed not only

by microscopy, but also in many biochemical studies. For example, distinct mapping of open and compact chromatin fibers to the G-band and C-band of chromosome was assayed by using hybridization to mitotic chromosomes and microarrays (GILBERT *et al.* 2004). In addition, reporter integration assays that randomly insert synthetic cassettes into the genome and monitor genomic integrations either by GFP output or high-throughput sequencing further confirmed the findings of compartment-wide regulation of gene expression (GIERMAN *et al.* 2007; AKHTAR *et al.* 2013).

Chromosome compartments can be qualitatively and quantitatively measured using Hi-C. Two types of structures called compartment A/B (representing active/inactive domains, respectively) were identified from Hi-C data by principle component analysis (PCA) (**Figure 1.8C**), where compartment A correlates with active epigenomic marks such as DNase I accessible regions, while compartment B relates to dense chromatin (LIEBERMAN-AIDEN *et al.* 2009). In addition to the discrete chromatin organization observed by electron microscopy (EM) imaging in different cell types, Hi-C mapping across a course of mammalian developmental and lineage specifications indicated that nearly 40% of compartment A/B switches throughout the genome (DIXON *et al.* 2015). These switches cooccur with transcriptional reprogramming during development, suggesting that chromosome compartments associate with gene activation/repression in distinct cell lineages. Whereas transcription activation is capable of inducing gene repositioning toward the nuclear interior as well as

compartment swapping, gene relocation can be also triggered by recruitment of an acidic peptide that decompacts chromatin without affecting transcription, indicating that both transcription and chromatin remodeling contribute to chromatin reorganization (THERIZOLS *et al.* 2014). However, a recent study indicated that although targeted genomic loci can be repositioned by recruitment of factors such as NANOG, SUV39H1, or EZH2, compartment switching was largely uncoupled from transcriptional changes, and histone modifications per se are not sufficient for repositioning (WIJCHERS *et al.* 2016). Taken together, the effects of *cis*- and *trans*- factors on chromatin compartments and transcription remains to be further clarified.

Interestingly, in addition to the transcription-related functions of large-scale chromosome compartments, comparing the nuclear architecture of rod photoreceptor cells between nocturnal and diurnal mammals found the inverted pattern, where heterochromatin localizes in the center of nucleus and acts as a collecting lens to channel light more efficiently. By contrast, the conventional architecture prevailing in most eukaryotic cells confers more flexible chromosome arrangements (SOLOVEI *et al.* 2009). These findings suggest that chromatin organization may provide an extra layer of function linking to many more biological systems.

Genome-wide chromosome conformation mapping and physical simulation suggest a fractal globule structure that enables the highest level of compaction while preserving the capacity to fold and unfold any genomic locus

(LIEBERMAN-AIDEN *et al.* 2009). While a few pioneer studies strategically removed key architectural proteins such as Cohesin and CTCF *in vivo*, their effects on the scales of chromatin structures differed from each other (also discussed in the next section) (SEITAN *et al.* 2013; NORA *et al.* 2016; SCHWARZER *et al.* 2016). The controversial findings strongly presume that, besides the *trans* factors mentioned, histone modifications, nucleosome positioning, and primary DNA sequences may intrinsically drive the formation of chromosomes.

Chromatin domains, boundaries, and loops

The compaction and organization of the genome into a physical genome have wide-ranging consequences for genomic function. Mounting evidence provided by single-locus studies at the globin and Hox clusters imply that chromatin could be organized as regulatory units in response to transcriptional regulation during development (DOSTIE *et al.* 2006; SIMONIS *et al.* 2006; BAU *et al.* 2011; NOORDERMEER *et al.* 2011; WILLIAMSON *et al.* 2014). Recent 3C-based techniques (e.g., 5C and Hi-C) with higher sequencing depth have discovered widespread chromatin domains (**Figure 1.8C**) (DIXON *et al.* 2012; NORA *et al.* 2012), which were further verified by super-resolution imaging combined with multiplexed error robust fluorescence *in situ* hybridization (MERFISH) (WANG *et al.* 2016). These techniques pioneered the discovery that chromosomes are partitioned into self-interacting domains, often called topologically associating domains (TADs) or chromosomally interacting domains (CIDs). TADs and CIDs typically manifest as

contiguous squares along the diagonal of contact matrices in 5C or Hi-C data, where regions within the same domain preferentially interact with each other over with regions located in adjacent domains (**Figure 1.9**).

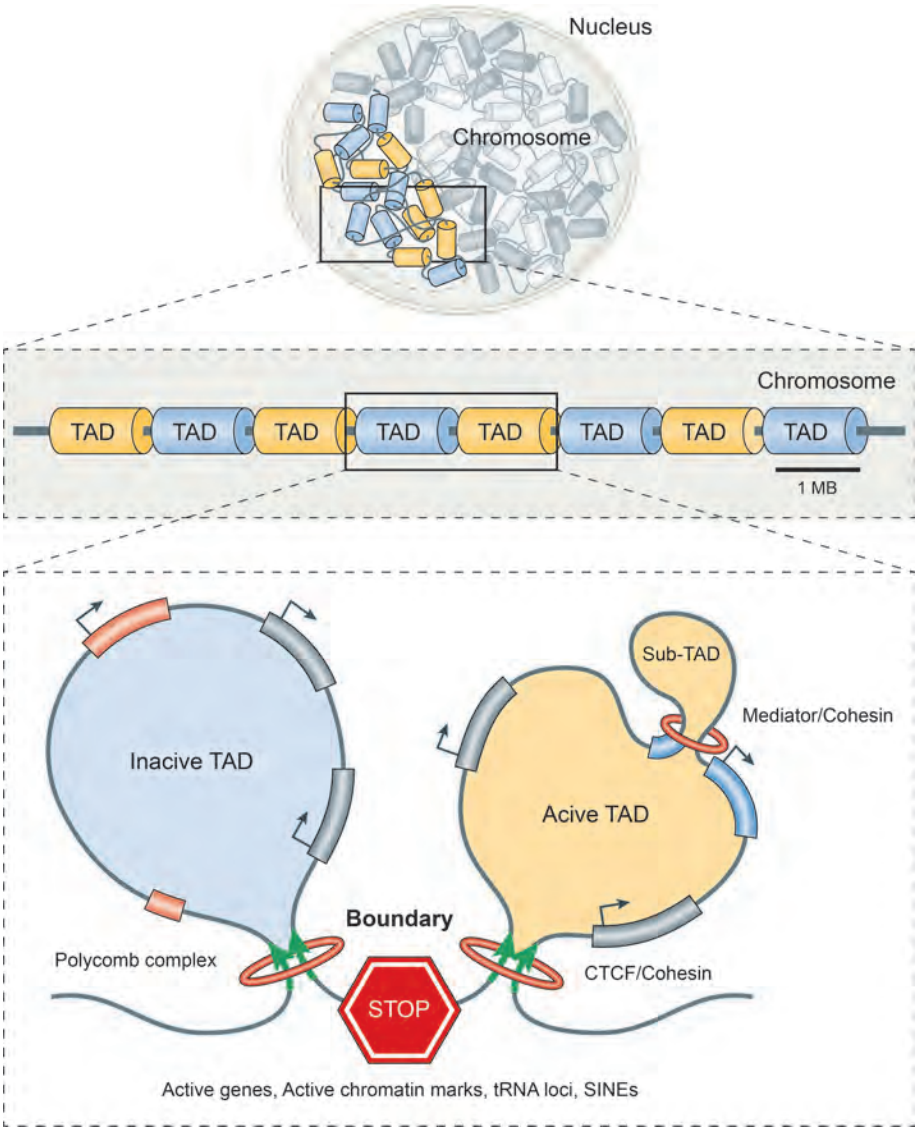


Figure 1.9. TADs, boundaries, and loops.

Chromosomes are subdivided into consecutive TADs, which limit the interactions crossing the boundaries from one TAD to another. Boundaries are enriched with active genes and chromatin marks, tRNA loci, and retrotransposons, as well as the architectural proteins like CTCF, Cohesin, or Mediator. Boundaries bound by CTCF and

Cohesin are responsible for the constitute TADs, while Mediator and Cohesin facilitate the formation of sub-TADs corresponding to the regulatory interactions (e.g., enhancer-promoter) during differentiation. Recent studies additionally identified domains/loops mediated by Polycomb complex, linking promoter areas of silenced genes. The formation of TADs/loops can be partially explained by the loop extrusion model – Cohesin complex forms progressive large loop/domain (blue or yellow domains/loops), while the looping process is stalled at the convergent CTCF-binding sequences (green arrows) bound by CTCF. Note that not all the TADs/loops can be simply explained by the loop extrusion model.

- Reprinted and adapted with permission from Nature Publishing Group: (KRIJGER AND DE LAAT 2016).

TADs have been identified in many cell types and seemed to be evolutionary conserved across species, including human (*Homo sapiens*), dog (*Canis familiaris*), mouse (*Mus musculus*), and macaque (*Macaca mulatta*) (DIXON *et al.* 2012; NORA *et al.* 2012; VIETRI RUDAN *et al.* 2015) as well in the model organisms fly (*Drosophila melanogaster*) and bacteria (*Caulobacter crescentus*)(HOU *et al.* 2012; SEXTON *et al.* 2012; LE *et al.* 2013; LE AND LAUB 2016). However, these domains are absent in organisms with small genomes, such as yeasts and plant (*Arabidopsis thaliana*)(DUAN *et al.* 2010; FENG *et al.* 2014; GROB *et al.* 2014; WANG *et al.* 2015), as well as the autosomes in worm (*Caenorhabditis elegans*)(CRANE *et al.* 2015). As a result, domains contain a hierarchical organization which can be subdivided into smaller domains. How these domains are identified and classified heavily depends on the genome size of the organism, the resolution of the experimental output, and the analysis approaches; that is, smaller genome organisms (such as yeast) may need

extremely high resolution data to visualize its chromatin domains. It is worth noting that the initial study of domains reported the median size of mammalian TADs to be ~880 kb, but subsequent Hi-C data with higher resolution reported a smaller size of ~200 kb, suggesting a great deal of uncertainty in defining self-interacting domains (DIXON *et al.* 2012; RAO *et al.* 2014). It also suggests that TADs previously identified in fly (~100 kb), the X chromosome of worm (~ 1 Mb), and even in bacteria (~ 50 kb) should be revisited with improved sequencing depth or resolution.

It has been suggested that the properties of chromatin domains closely associate with their primary chromatin features (such as histone modifications and nucleosome occupancy), which generally reflects transcriptional activity or replication timing. For example, a subset of chromatin domains called the lamina associating domains (LADs) is classified by discrete regions with low gene-expression and high repressive histone marks (H3K27me3 and H3K9me3) (GUELEN *et al.* 2008; REDDY *et al.* 2008; KIND *et al.* 2013; KIND *et al.* 2015). Cytologically, these regions represent as heterochromatic domains surrounding the nuclear periphery. Artificial tethering genomic loci to LADs or integrating synthetic sequencing within LADs leads to transcriptional inactivation (REDDY *et al.* 2008). Additionally, mammalian genomes are segmented into replication domains (RDs) based on replication timing, which are highly correlated with the properties of chromosome compartments (TAKEBAYASHI *et al.* 2012; POPE *et al.* 2014). Although the cell-type specific RDs occur at the TAD level, early RDs

usually co-localize with active/open compartments and the late RDs reside in inactive/compact compartments. Despite these strong correlations, the cause or consequence relationship between chromatin and nuclear transactions still remains elusive. For example, TADs can function as regulatory units in response to hormone-induced transcriptional reprogramming (LE DILY *et al.* 2014), as well as associate with many transcription-related developmental diseases (e.g. sex reversal, cook syndrome, and cancer-related) (CHANDRA *et al.* 2015; LUPIANEZ *et al.* 2015; FRANKE *et al.* 2016). In addition, transcriptional activity seems to be co-regulated within an individual TAD, as seen by enhancer-insertion assays in mouse embryos (SYMMONS *et al.* 2014). However, at least in X chromosome, G9a-/- (H3K9 methyltransferase) or Eed-/- (H3K27 methyltransferase) that disrupt the depositions of heterochromatic marks does not lead to any significant change in TAD organization (NORA *et al.* 2012), which suggests a more complex regulatory network between chromatin structures and transcription or TAD formation being upstream of histone modifications and transcriptional regulations.

Chromatin domains are spatially demarcated by domain boundaries, which are defined by calling the lowest local cross-interactions in a given size scanning window (DIXON *et al.* 2012). The boundaries are enriched for architectural factors such as CTCF, Cohesin, topoisomerases, active genes (nascent transcripts) and histone marks (H3K4me3, H3K36me3), tDNA loci, and short interspersed element (SINE) retrotransposons (**Figure 1.9**) (DIXON *et al.* 2012; PHILLIPS-CREMINS *et al.* 2013; UUSKULA-REIMAND *et al.* 2016). Genome-editing the CTCF

binding site by either inversion or deletion severely disrupts its insulation activity (GUO *et al.* 2015), resulting in the spreading of active domains to repressive domains on Hox loci (NARENDRA *et al.* 2015), or ectopic expression of pathogenic genes on limb malformations (LUPIANEZ *et al.* 2015). These results strongly suggest a functional importance to boundaries in insulating chromatin for proper gene regulations. In addition to insulation activity, nearly 40% of mammalian TAD borders can form chromatin loops or loop domains (HANDOKO *et al.* 2011; PHILLIPS-CREMINS *et al.* 2013; RAO *et al.* 2014; DE WIT *et al.* 2015), which is further supported by the loop extrusion model where boundary-limited looping in a dynamic fashion is sufficient to predict TAD formation (**Figure 1.9**) (SANBORN *et al.* 2015; FUDENBERG *et al.* 2016).

Various combinations of architectural elements on the boundary determine the properties of boundaries, loops, and domains, whose mechanisms have been extensively investigated during the processes of differentiation and reprogramming (**Figure 1.9**) (KAGEY *et al.* 2010; PHILLIPS-CREMINS *et al.* 2013; BEAGAN *et al.* 2016; KRIJGER *et al.* 2016). In brief, CTCF/Cohesin anchors the long-range interactions (> 1 – 2 Mb) that form the constitutive boundary-boundary interactions/domains in both pluripotent and differentiated cells. In contrast, Mediator/Cohesin bridge interactions in the shorter-range (100 kb – 500 kb), reflecting cell-type specific enhancer-promoter or promoter-promoter interactions (also called sub-TADs), which has been broadly mapped via ChIA-PET (on Pol2, CTCF, or Cohesin) and Capture Hi-C in many cell types (HANDOKO *et al.* 2011; LI

et al. 2012; SANDHU *et al.* 2012; DEMARE *et al.* 2013; ZHANG *et al.* 2013; MIFSUD *et al.* 2015; SCHOENFELDER *et al.* 2015). These results indicated that transcription clusters (often called transcription hubs) are non-randomly organized in the nucleus, and usually co-regulated by a group of key elements. In addition, CTCF/Cohesin is capable of looping out 'super-enhancer domains' (enriched for Pol2, Mediator, and H3K27Ac) and 'polycomb domains' (enriched for H3K27me3) to physically constrain their activity from neighboring chromatin (DOWEN *et al.* 2014). However, the molecular basis of boundary activity has not yet been resolved, as the results of depleting CTCF or Cohesin contradict one another. For example, depletion of Cohesin (Rad21) only caused marginal changes in TADs and loops (SEITAN *et al.* 2013; SOFUEVA *et al.* 2013; ZUIN *et al.* 2014); on the contrary, deletion of Cohesin-loading complex (NipbI) led to dramatic chromatin reorganization, in which TADs and loops were diminished globally but compartments were preserved and even reinforced (SCHWARZER *et al.* 2016). Similarly, one study indicated a drastic loss of chromatin structures upon removal of CTCF (NORA *et al.* 2016), while another one suggested CTCF-independent maintenance of chromatin folding (KUBO *et al.* 2017). These contrary results could result from differential levels of depletion of the target protein, or differences in experimental protocols or analysis approaches, and thus need future studies.

Chromatin organization in Drosophila

Chromatin domains in flies resemble the properties and functions in mammals, which can be classified by the “five-color” chromatin types including active chromatin, Polycomb-mediated domains, HP1-mediated domains, and lamina-associating domains (FILION *et al.* 2010). Inactive domains are more compacted and confined within their chromosomal territories, while active domains are flexible and more likely to form long-range interactions with other active regions (HOU *et al.* 2012; SEXTON *et al.* 2012; EAGEN *et al.* 2015; ULIANOV *et al.* 2016). This agrees with consistent results from super-resolution imaging showing that polycomb chromatin has more compacted folding and less interdomain overlapping than active domains (BOETTIGER *et al.* 2016). Boundaries in flies also are comparable to those in mammals, with enrichment of active transcription and insulators like BEAF, CTCF, CP190, and Su(Hw) at inter-TAD regions, even though transcription plays a more dominant role in the boundary activity in flies (HOU *et al.* 2012; SEXTON *et al.* 2012; LIANG *et al.* 2014; VOGELMANN *et al.* 2014; ULIANOV *et al.* 2016). Interestingly, transcriptional repression upon heat-shock stress results in the dramatic rearrangement of the chromatin organization in flies (LI *et al.* 2015). These reorganizations are controlled by the redistribution of architectural proteins from the borders of TADs to inner regions, in which the polycomb complex facilitates enhancer-promoter interactions between silenced genes. In general, chromatin looping is not a widespread feature in flies. Recent Hi-C experiments with deeper sequencing depth identified ~120 chromatin loops whose formation are mediated by PRC1/H3K27me3 (**Figure 1.9**) (EAGEN *et al.*

2017). Interestingly, TADs identified in diploid cells were found corresponding to the banding pattern in polytene chromosomes (shown in prior microscopic studies), while polytene puffs co-localize to promoters and regulatory regions, where the chromatin fibers are fully extended as a 'beads-on-a-string' structure in inter-TAD regions (EAGEN *et al.* 2015). These findings suggest that chromatin domains, boundaries, and loops are highly conserved in drosophila, and tightly associate with many regulatory events.

Chromatin organization in Arabidopsis

In plants such as *Arabidopsis thaliana*, the existence of TADs or TAD-like domains still remains obscure, in part owing to insufficient resolution of current techniques (FENG *et al.* 2014; GROB *et al.* 2014; WANG *et al.* 2015; LIU *et al.* 2016). Besides heterochromatic islands such as TEL-TEL and CEN-CEN clusters and KNOTs (the cluster of transposable elements), two independent studies were only able to find a subset of relatively small interacting regions scattered around the genome, in which specifically enriched for H3K27me3 and H3K9me2 (FENG *et al.* 2014; WANG *et al.* 2015). Nevertheless, a recent finding analyzed Hi-C data at sub-kilobase segmentation and argued for the existence of H3K27me3-associated chromatin loops and active/inactive gene looping structures in *Arabidopsis* chromatin (LIU *et al.* 2016). Therefore, whether self-associating domains or loops exist in plants and what functions they being responsible for need to be further characterized in future.

Chromatin organization in yeasts and gene loops

Rather than having chromosome territories or compartments, yeast organizes chromosomes in a ‘Rabl-like’ conformation including three conspicuous features (**Figure 1.10**): 1) the centromeres are anchored at one end of nuclear periphery together with the spindle pole body (CEN-CEN cluster); 2) the chromosome arms are expelled away from centromeric regions (CEN-Arm avoidance); and 3) the telomeres on the equivalent length of chromosome arms form the scattered clusters on another end of the peripheric region (TEL-TEL clusters) (DUAN *et al.* 2010; ZIMMER AND FABRE 2011; TADDEI AND GASSER 2012). The ‘Rabl-like’ structure on Chromosome III also explains how yeast structurally regulates the switching of its mating-type loci (MIELE *et al.* 2009; BELTON *et al.* 2015). Interestingly, yeast cells can be triggered to enter into quiescence stage (G0) upon suffering starvation, which leads to global chromosomal compaction accompanied with loss of centromere interactions and an increase of telomere clustering (GUIDI *et al.* 2015; RUTLEDGE *et al.* 2015). Beside this large-scale organization, recent high-resolution Hi-C work revealed ‘Cohesin-delimited globules’ in the fission yeast *Schizosaccharomyces pombe*, although no similar level or type of chromatin domain has been identified yet in budding yeast (DUAN *et al.* 2010; MIZUGUCHI *et al.* 2014).

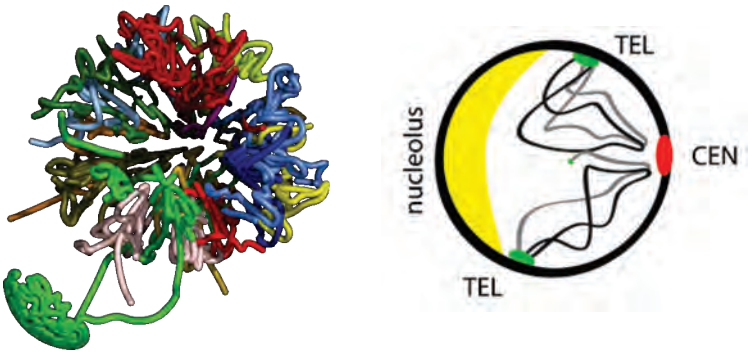


Figure 1.10. Yeast chromosome organization.

(*Left panel*) 3D modeling of chromosome architecture by using yeast Hi-C. The center of nucleus is showing the centromere cluster and the telomeres are scattered at the nuclear periphery. The outside globule represents the rDNA loci at ChrXII. (*Right panel*) Schematic of classic 'Rabl-like' chromosome architecture in yeast, showing the focal centromere cluster, the scattered telomere clusters, and chromosome arms expelled from the centromeric region.

- Left panel: Reprinted and adapted with permission from Nature Publishing Group: (DUAN et al. 2010)
- Right panel: Reprinted and adapted with permission from GSA: (TADDEI AND GASSER 2012).

Gene looping structures, primarily identified in yeast, have been proposed to facilitate overall Pol2 transcription including initiation, termination, and reassembling of transcription machinery. Many factors such as general transcription factors (TFIIB and TFIIH), activators, Mediators, cleavage factors, and poly-A and intron sequences all have been reported to be involved in the formation of gene loops (O'SULLIVAN *et al.* 2004; ANSARI AND HAMPSEY 2005; SINGH AND HAMPSEY 2007; TAN-WONG *et al.* 2009; MUKUNDAN AND ANSARI 2011; AL HUSINI *et al.* 2013; MUKUNDAN AND ANSARI 2013). Recent study reported that gene looping structures can enforce transcriptional directionality, which minimize

cryptic transcription occurred at promoter regions (TAN-WONG *et al.* 2012). Either mutation of 'looping factor' SSU72 or disruption of gene loops by removal of poly-A site increases a wide-ranging synthesis of promoter-associated divergent ncRNAs. In contrast to promoting transcriptional activity, one study suggested that ISW2/UME6-mediated gene loops associate with transcriptional repression (YADON *et al.* 2013). However, all these gene looping evidence are from single-locus studies by low-resolution 3C. Further investigation with genome-wide or high-resolution approaches is needed to determine whether the structure exists universally.

Unsolved problems

As mentioned above, understanding higher-order chromatin structure has been greatly facilitated by the 3C family of techniques, which discern contact frequency between genomic loci based on isolation of DNA fragments that crosslink to one another *in vivo*. While many factors impact the effective resolution of a 3C/Hi-C dataset, including sequencing depth and library complexity, a fundamental limit to genomic resolution is the size of the fragments generated before physical interactions are captured via ligation. Since the majority of 3C-based experiments rely on restriction enzymes for fragmentation of the genome – resulting in genomic fragments that are both long relative to the nucleosome, and heterogeneously spaced along the genome – current Hi-C datasets are limited to ~1 kb resolution at best. Thus, our present understanding of chromatin structure

has a blind spot, with ChIP-Seq, MNase-Seq, and ChIP-exo methodologies providing information over a ~1-150 bp length scale, and Hi-C typically providing information on the >1-4 kB length scale. These techniques thus leave the scale relevant to secondary structures such as the 30 nm fiber or yeast gene loops (on the order of ~2-10 nucleosomes) inaccessible to current methods for analyzing chromosome structure.

CHAPTER II

Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C

Abstract

We describe a Hi-C based method, Micro-C, in which micrococcal nuclease is used instead of restriction enzymes to fragment chromatin, enabling nucleosome resolution chromosome folding maps. Analysis of Micro-C maps for budding yeast reveals abundant self-associating domains similar to those reported in other species, but not previously observed in yeast. These structures, far shorter than topologically-associating domains in mammals, typically encompass one to five genes in yeast. Strong boundaries between self-associating domains occur at promoters of highly transcribed genes and regions of rapid histone turnover that are typically bound by the RSC chromatin-remodeling complex. Investigation of chromosome folding in mutants confirms roles for RSC, “gene looping” factor Ssu72, Mediator, H3K56 acetyltransferase Rtt109, and the N-terminal tail of H4 in folding of the yeast genome. This approach provides detailed structural maps of a eukaryotic genome, and our findings provide insights into the machinery underlying chromosome compaction.

Introduction

Eukaryotic genomes are packaged into chromatin via a hierarchical series of folding steps. A great deal is known about the first level of chromatin compaction, as several crystal structures exist of the repeating subunit – the nucleosome – and genome-wide mapping studies have illuminated nucleosome positions and histone modifications across the genome for an ever-increasing number of organisms (RANDO 2007; ZHANG AND PUGH 2011; HUGHES AND RANDO 2014). In contrast to the "primary structure" of chromatin, less is known about higher-order chromatin architecture. The next level of compaction is commonly thought to be the 30 nm fiber, which is readily observed by electron microscopy in vitro, but whose existence in vivo remains controversial (TREMETHICK 2007; FUSSNER *et al.* 2011; MAESHIMA *et al.* 2014). The structure of a 30 nm fiber is hotly debated, with major models being solenoid and zigzag paths of the beads-on-a-string (DORIGO *et al.* 2004; TREMETHICK 2007; GHIRLANDO AND FELSENFELD 2008; ROUTH *et al.* 2008; SONG *et al.* 2014), as well as more recent polymorphic fiber models that incorporate variability in nucleosome repeat length (COLLEPARDO-GUEVARA AND SCHLICK 2014). Moreover, mounting evidence suggests that 30 nm fiber may only occur in vitro due to the high dilution of chromatin fibers used in such studies – in dilute solution in vitro a given nucleosome will only have access to other nucleosomes on the same DNA fragment, while in the "sea of nucleosomes" in the nucleus many additional nucleosomes are available in trans for internucleosomal interactions (MCDOWALL *et al.* 1986; NISHINO *et al.* 2012). Beyond the 30 nm fiber, multiple additional levels of organization have been

described, with prominent examples including gene loops (O'SULLIVAN *et al.* 2004; ANSARI AND HAMPSEY 2005), enhancer-promoter loops (SANYAL *et al.* 2012), “topologically-associating domains”/”chromosomally-interacting domains” (TADs/CIDs) (DIXON *et al.* 2012; NORA *et al.* 2012; SEXTON *et al.* 2012; LE *et al.* 2013), lamina-associated domains (LADs) (PICKERSGILL *et al.* 2006), and megabase-scale active and repressed chromatin compartments (LIEBERMAN-AIDEN *et al.* 2009; GROB *et al.* 2014). The 3-dimensional path of chromatin has been implicated in a large number of biological processes, as for example gene loops are proposed to enforce promoter directionality in yeast (TAN-WONG *et al.* 2012), TADs correspond to regulatory domains in mammals (SYMMONS *et al.* 2014), and LADs are correlated with gene silencing during development (PICKERSGILL *et al.* 2006).

Understanding higher-order chromatin structure has been greatly facilitated by the 3C family of techniques (such as Hi-C), which assay contact frequency between genomic loci based on isolation of DNA fragments that crosslink to one another in vivo (DEKKER *et al.* 2002). However, these techniques currently suffer from suboptimal resolution, as they rely on restriction digestion of the genome, typically yielding ~4 kb average fragment size. Even with 4-cutter restriction enzymes, the heterogeneous distribution of restriction enzyme target sequences across the genome makes the resolution somewhat variable between individual loci of interest, and partial digestion still limits resolution to around 1 kb at best. Thus, our present understanding of chromatin structure has a “blind

spot", with ChIP-Seq, MNase-Seq, and ChIP-exo methodologies providing information over the ~1-150 bp length scale, and Hi-C typically providing information on the >1-4 kB length scale. This leaves the length scale relevant to secondary structures such as 30 nm fiber or yeast gene loops – on the order of ~2-10 nucleosomes – inaccessible to current methods for analyzing chromosome structure.

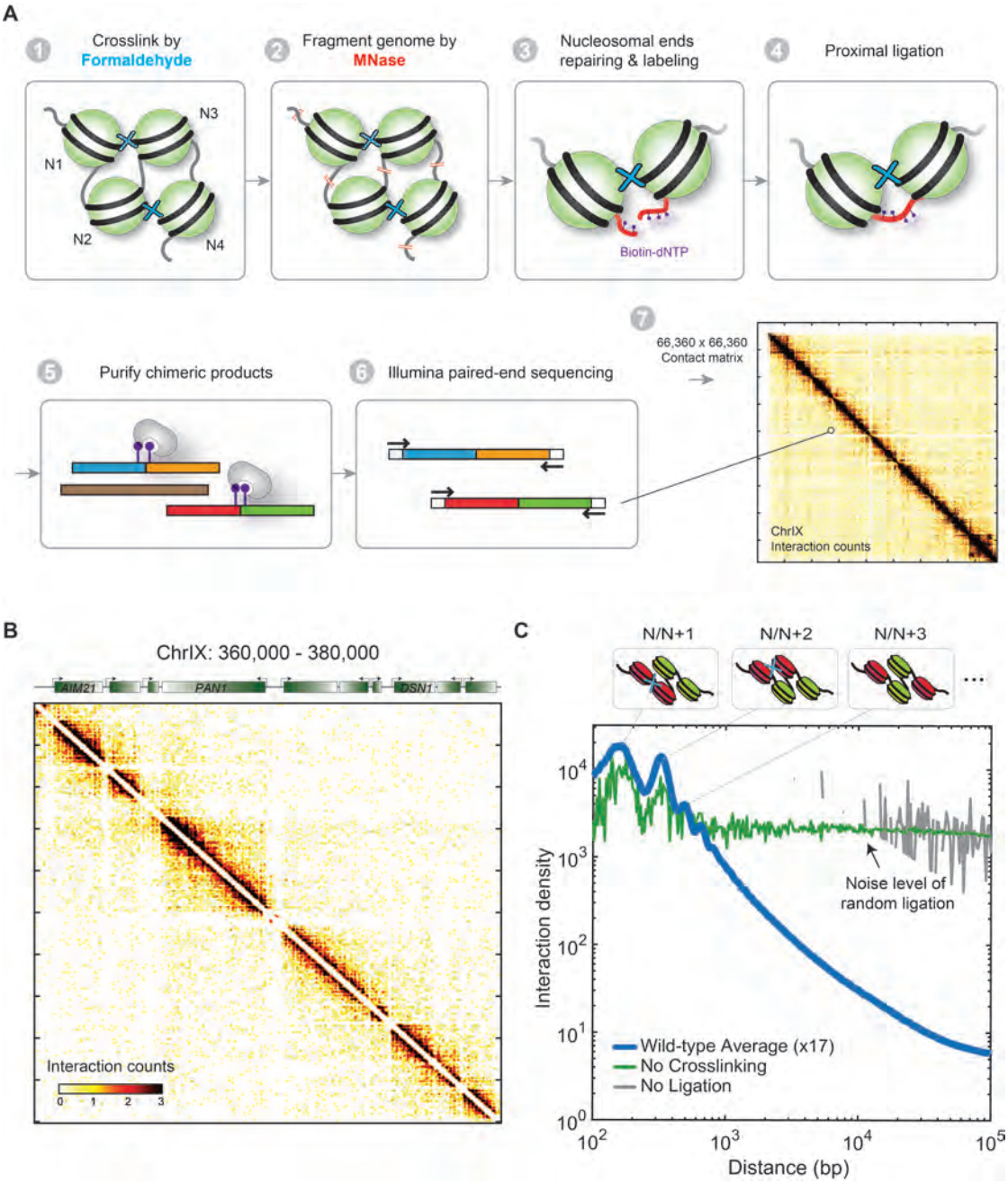
Here, we describe a Hi-C-based method – "Micro-C" – in which chromatin is fragmented into mononucleosomes using micrococcal nuclease, thus enabling nucleosome-resolution maps of chromosome folding. We generated high-coverage Micro-C maps for the budding yeast *S. cerevisiae*, finding abundant self-associating domains typically spanning 1-5 genes. Strong boundaries between self-associating domains occur at promoters of highly-transcribed genes and regions of rapid histone turnover, and are typically bound by the RSC ATP-dependent chromatin remodeling complex and by the cohesin loading complex. Finally, we investigate chromosome folding in detail in 14 mutants, confirming roles for RSC and Ssu72 in chromosome folding, and furthermore finding key roles for Mediator, the histone H4 N-terminal tail, and the H3K56 acetyltransferase Rtt109 in folding of the yeast genome. This approach thus enables analysis of chromosome folding at the resolution of chromatin's repeating subunit – the nucleosome – and will enable future investigations into chromosome folding to leverage the powerful genetic tools available in the yeast model system.

Result

A nucleosome resolution chromosome folding assay

The resolution gap between 1D chromosome mapping assays (~1-200 bp resolution) and 3D chromosome folding assays (>1 kb resolution) lead us to develop a Hi-C protocol – termed “Micro-C” for MICROcoccal nuclease chromosome Conformation assay – in which chromatin fragmentation is achieved by MNase digestion, yielding mononucleosomes. This protocol is based on the Hi-C protocol (LIEBERMAN-AIDEN *et al.* 2009), with key alterations being the MNase digestion step, subsequent mononucleosomal end repair, and a modified two-step method for specifically purifying ligation products (**Figure 2.1A, Methods**). After purification of ligation products between mononucleosomes, paired-end deep sequencing is used to characterize the ligation products. For some analyses, each pair mate is assigned to one of the 66,360 nucleosomes in budding yeast (WEINER *et al.* 2010) to yield a 66,360 X 66,360 nucleosome-nucleosome interaction matrix (**Figure 2.1B**). Key technical controls include 1) sequencing of unligated samples, 2) ligation of MNase digestion reactions of uncrosslinked yeast, and 3) mixing of crosslinked chromatin from two yeast species (*S. cerevisiae* and *K. lactis*) prior to ligation to determine the rate of ligation between uncrosslinked molecules (**Figure 2.1C-D**). The post-crosslinking interspecies mixing experiment reveals that < 5% of all interactions are spurious ligations at the dilution used.

Our technique provides an overlapping but nonidentical view of chromosome folding to restriction enzyme-based methods. Visual inspection of a prior low-resolution chromosome folding map for budding yeast (DUAN *et al.* 2010) confirms the substantially higher resolution of our assay (**Figure 2.1B-D**). However, Micro-C poorly captures known long-distance interactions in yeast: while we do recover preferential interactions between short chromosomes and a modest signal for telomere-telomere interactions, these interactions are relatively weak, and centromere-centromere interactions are not observed (**Figure 2.1E-F**). Micro-C thus serves as a complementary method to traditional 3C and Hi-C methods, being particularly well suited to short-range analysis of nucleosome fiber folding, which is invisible to restriction enzyme-based 3C/Hi-C assays.


A
1 Crosslink by Formaldehyde
2 Fragment genome by MNase
3 Nucleosomal ends repairing & labeling
4 Proximal ligation
N1
N2
N3
N4
Biotin-dNTP
5 Purify chimeric products
6 Illumina paired-end sequencing
7 66,360 x 66,360 Contact matrix
ChrIX Interaction counts
B
ChrIX: 360,000 - 380,000
AIM21
PAN1
DSN1
Interaction counts
0 1 2 3
C
N/N+1
N/N+2
N/N+3
Interaction density
Noise level of random ligation
Wild-type Average (x17)
No Crosslinking
No Ligation
Distance (bp)

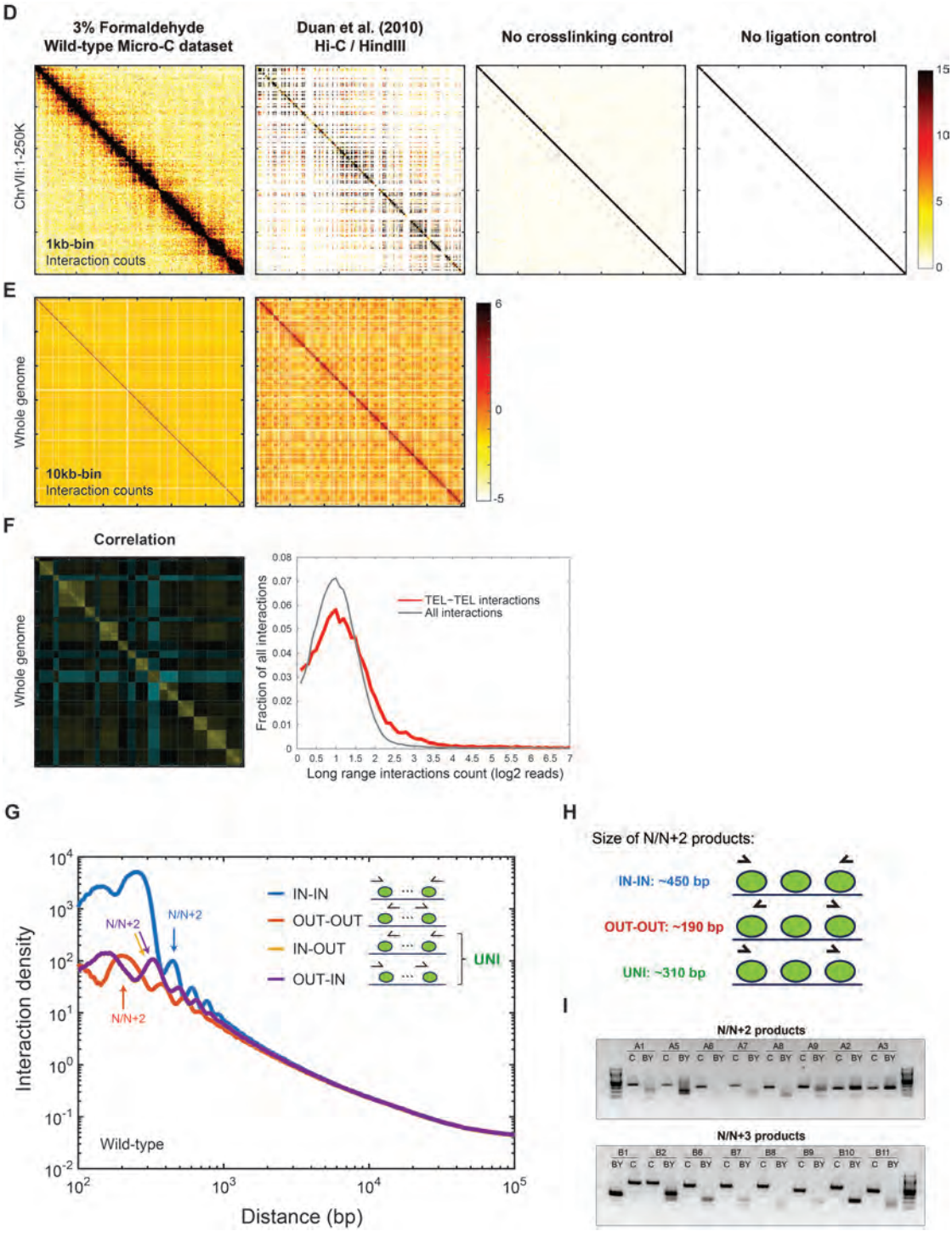


Figure 2.1. Nucleosome-resolution chromosome folding maps.

(**A**) Overview of the Micro-C method. Bottom right panel shows Micro-C data for yeast chromosome IX. (**B**) Zoom-in on a 20 kb X 20 kb submatrix from chromosome 9

(360,000-380,000), with Micro-C interactions represented in white-yellow-red-black heatmap showing the interaction intensity for a given pair of loci. (**C**) Decay of internucleosomal interactions with distance. Distances along the x axis are provided in units of nucleosomes – first data point represents ligation between adjacent (N/N+1) nucleosomes, with data out to ~N/N+60 (100 kb) products shown. Y axis shows square root of the number of ligation products, normalized to parts per million (for interactions out to 100 kb) for each dataset. Both axes are shown in $\log_{10}$ scale. The plot only includes the "UNI" interactions for eliminating the bias introduced by undigested di-nucloesomal DNA (**Figure 2.1G**). Data for the average of 17 wild-type replicates, and for no crosslinking and no ligation control datasets, are indicated. Schematics illustrate nucleosomes contributing to N/N+1, N/N+2, and N/N+3 ligation products, using a tetranucleosome cartoon for illustration. (**D**) Micro-C signal is dependent on crosslinking and ligation steps. Micro-C data are shown for merged wild-type BY4741 replicates (first panel), and for control Micro-C reactions carried out without formaldehyde crosslinking of yeast (third panel), or with a mock ligation step (fourth panel). Data are binned by 1 kb in this view, and are shown for a 1 – 250 kb region zoom-in for chromosome 6. Comparison at a finer resolution with 1kb binning shows the sparseness of restriction-based assay (second panel) compared to Micro-C. (**E**) Comparison of Micro-C with restriction-based mapping (DUAN *et al.* 2010). Both raw datasets were processed using our pipeline (to avoid processing bias) and interaction counts for 10 kb bins are shown. At this scale, the main differences between the two maps are centromere-centromere and telomere-telomere interactions that are pronounced in the restriction-based assay result and not easily apparent in Micro-C. (**F**) (Left panel) Correlation matrix for Micro-C data. For each 20 kb bin along the genome, the correlation to all other bins is shown in a heatmap (yellow = positive, blue = negative). Focusing on interactions between chromosomes, the shorter chromosomes in yeast (chromosomes 1, 3, 6, 9) tend to exhibit better correlations with one another than with the longer chromosomes. (Right panel**)** Recovery of telomere-telomere interactions in Micro-C data. Distributions show the fraction of all interactions observed for long range interactions among random genomic loci, or among telomere-proximal loci (<20 kb from a chromosome end). Subtelomeres interact with distal subtelomeric regions more often than do random pairs of distant genomic loci. In addition, chromosome-specific patterns of telomere-telomere

interactions were similar for Micro-C and for (DUAN *et al.* 2010) (not shown). (**G**) Decay of product abundance with distance is shown separately for Micro-C products (merged wild-type dataset) with read pairs facing towards one another in the yeast genome (IN-IN), and for the IN-OUT and OUT-OUT read orientations (UNI) – IN-OUT and OUT-IN curves are of course nearly identical. Each dataset is normalized to parts per million relative to all interactions out to 100 kb, and both axes are shown in log10 scale. Readily apparent here is a strong bias among IN-IN reads for undigested mononucleosomes and for N/N+1 products originating from undigested dinucleosomal DNA that is not eliminated by the biotin selection step (most likely due to incomplete exonuclease digestion after ligation – **Figure 2.1A**). Arrows show the location of N/N+2 ligation products for each curve. Importantly, for IN-OUT and OUT-OUT products, N/N+2 products are nearly as abundant as N/N+1 ligation products. (H) Schematic showing read pairs for N/N+2 ligation products in various read orientations, providing intuition for the read pair distances indicated with arrows in (G). (**I**) N/N+2 and N/N+3 IN-IN products generally do not result from intact tri- or tetra-nucleosomes being sequenced in Micro-C. Inward-facing PCR primers were designed for highly abundant Micro-C N/N+2 and N/N+3 ligation products. For each primer pair PCR was carried out on genomic DNA (C) or a BY4741 Micro-C library (BY). With two exceptions for N/N+2 (A2 and A3, both of which came from “fuzzy” regions of delocalized nucleosomes) the genomic product was ~150 or ~300 bp longer than the Micro-C library product, demonstrating that the Micro-C product resulted from ligation of digested mononucleosomes, and could not result from intact tri or tetranucleosomes. Nonetheless, because of the significant overabundance of IN-IN reads at short distances we exclude IN-IN ligation products for all downstream analyses.

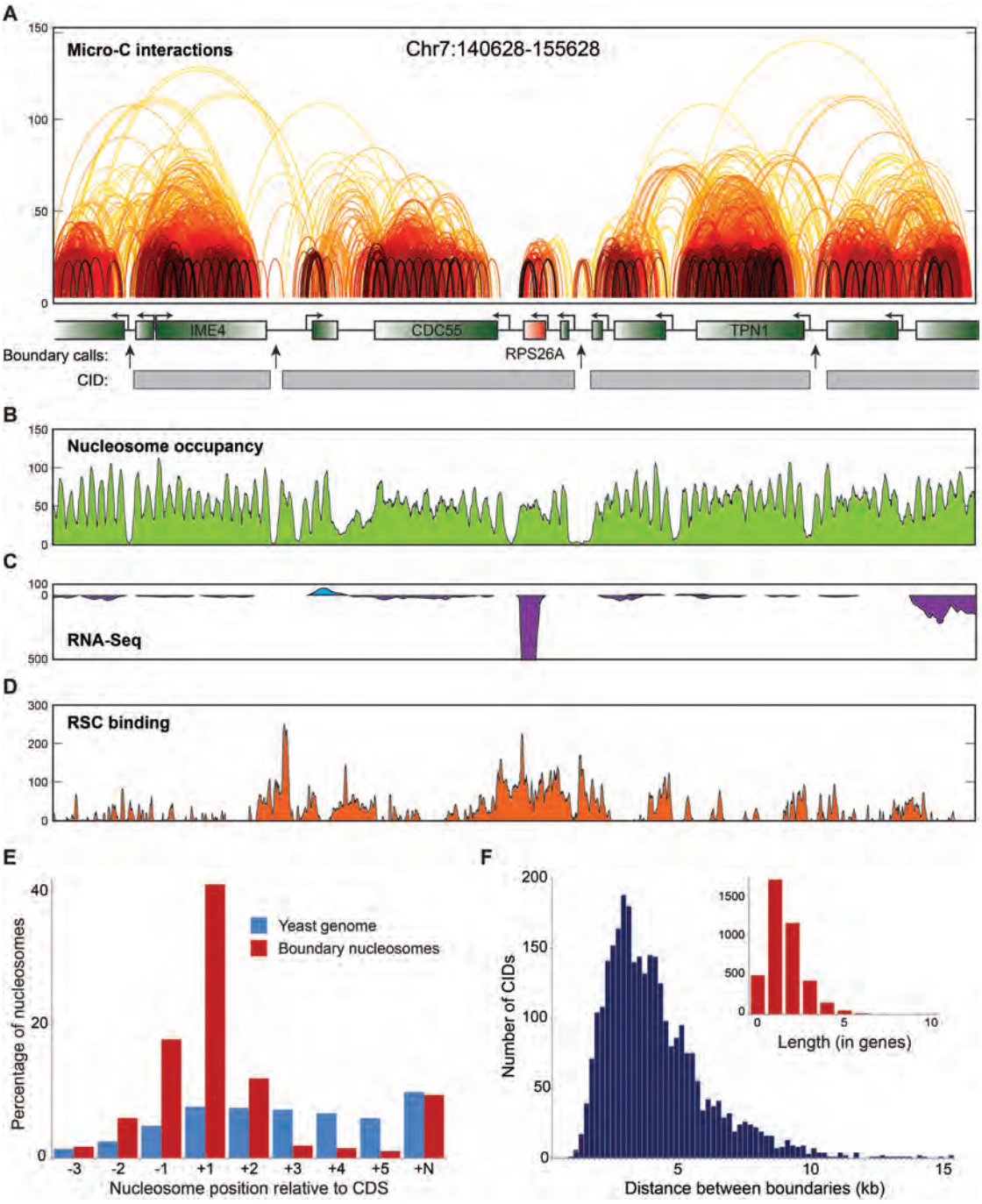
Promoter nucleosomes form boundaries between self-associating domains

As is typical of all chromatin interaction maps, the vast majority of Micro-C interactions occur close to the diagonal – nucleosomes tethered near one another in one dimension tend to contact each other in 3D space (**Figures 2.1B-C**). Nonetheless, there is significant variation in the density of interactions along

the diagonal (**Figure 2.1B**), with abundant ~4-50 nucleosome "boxes" of internucleosome interactions that have a clear relationship to gene structure. These boxes of internucleosomal interactions are similar to the "topologically-associated domains" (TADs) described in mammals (DIXON *et al.* 2012; NORA *et al.* 2012) and the chromosomal interaction domains (CIDs) described in the bacterium *C. crescentus* (LE AND LAUB 2016), which have also been observed in flies (SEXTON *et al.* 2012) but appear to be absent in *A. thaliana* (FENG *et al.* 2014) and were not previously observed in *S. cerevisiae* (DUAN *et al.* 2010). Here we will adopt the more general "CID" nomenclature. As observed in multiple organisms, these interaction domains exhibit a nested architecture, with two strong short range domains often merging into a larger domain via somewhat weaker interactions. We consistently observe CIDs across 21 biological replicate samples (**Figure 2.2I**) for *S. cerevisiae* including three separate "wild type" strain backgrounds (S288C, W303, and a S288C strain with the histone H3/H4-encoding genes relocated to a plasmid (DAI *et al.* 2008)), as well as in a somewhat distantly-related (last common ancestor ~150 mY) hemiascomycete yeast, *K. lactis*.

We systematically identified boundaries between CIDs by searching for locations that are strongly depleted of crossing interactions relative to the density of such interactions in the region (**Figures 2.2A-D**). Boundary calls were consistent between replicates of wild-type yeast (**Figure 2.2I**), and were not an artifact of MNase digestion level (**Figures 2.2J-L**). Globally, boundaries

separating compacted local domains from one another were strongly enriched for the nucleosome depleted regions (NDRs) that are a widespread feature (HUGHES AND RANDO 2014) of yeast promoters (**Figures 2.2B and E**). That said, it is clear that not all promoters or NDRs form boundaries – CIDs ranged from 0 to 8 genes in length, with 45% of CIDs encompassing two or more genes (**Figure 2.2F**). Interestingly, as CIDs in budding yeast typically encompass one to five genes, at ~2-10 kb they are 1-2 orders of magnitude shorter than mammalian TADs (~100 kb - 1 Mb) – the length of self-associating domains thus appears to be conserved when scaled by gene number, as opposed to sequence distance.


A
Micro-C interactions
Chr7:140628-155628
150
100
50
0
IME4
CDC55
RPS26A
TPN1
Boundary calls:
CID:
B
Nucleosome occupancy
150
100
50
0
C
100
0
RNA-Seq
500
D
RSC binding
300
200
100
0
E
Yeast genome
Boundary nucleosomes
Percentage of nucleosomes
40
20
0
-3
-2
-1
+1
+2
+3
+4
+5
+N
Nucleosome position relative to CDS
F
Number of CIDs
200
150
100
50
0
5
10
15
Distance between boundaries (kb)
1500
1000
500
0
0
5
10
Length (in genes)

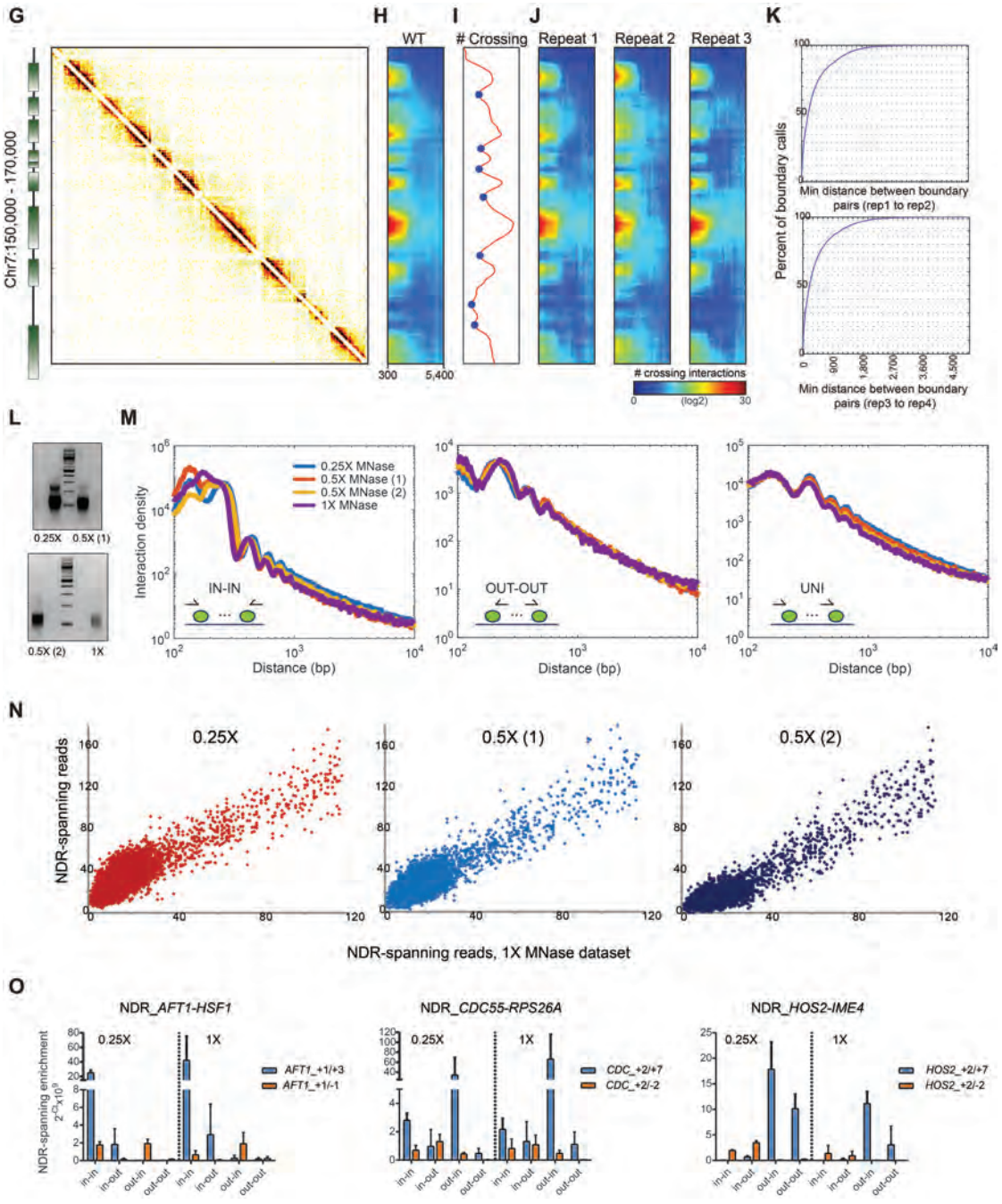


Figure 2.2. Properties of folding boundaries.

(**A-D**) Example of boundary identification. Data for a 15 kb locus, with arcs showing interactions between nucleosomes, colored as in **Figure 2.1B**. Interactions observed only once in the entire dataset have been removed for clarity. Gene annotations for this locus, and boundary calls shown in black arrows, are shown below panel (A). Emphasis

on *RPS26A* shows both the overall lack of local Micro-C interactions, as well as the unusually strong boundary activity associated with this highly-transcribed gene. Nucleosome positioning data (B), RNA-Seq data (C), and Sth1 ChIP-Seq enrichment (LOPEZ-SERRA *et al.* 2014) (D) are also shown for this locus to emphasize the correlation between RSC-enriched promoters and boundary activity. (**E**) Boundaries between CIDs occur at promoters. For each nucleosome position relative to a gene, the fraction of boundary nucleosomes, or of all nucleosomes genome-wide, is shown on the y axis. As boundaries as defined here fall between adjacent nucleosomes, we show data here for the *downstream* boundary nucleosome, relative to underlying gene orientation – upstream nucleosomes are correspondingly enriched for -1 nucleosomes (and +N nucleosomes). (**F**) Length distribution of CIDs. Distribution of distances between boundary nucleosomes is plotted in blue using base pairs for the x axis, and in the inset using gene count as the scale. (**G**) A typical interaction matrix for a 20 kb X 20 kb genomic region, as in **Figure 2.1B**. (**H**) Within the matrix shown in (A), number of interactions crossing a specific position (y axis) are shown for varying interaction distances (x axis) in a blue-to-red $\log_2$-scaled heatmap. (**I**) All reads for domain sizes from 500 – 10,000 bp crossing a given position are summed, and the local minimum of this vector is identified. (**J**) Data, shown as in (H), for three independent wild-type replicate datasets. (**K**) Consistent boundary calls between BY4741 replicate datasets. For four individual replicates with >30 million reads, Micro-C boundaries were called for each individual dataset. Plots show the fraction of boundaries called in one replicate that were within a given distance (x axis) of a boundary called in the other replicate. For each pairwise comparison, ~85-90% of all boundary calls were located within 1 kb of one another. Moreover, inspection of those boundary calls that disagreed revealed that most were found to occur at relatively long boundaries such as *RPS26A* in **Figures 2.2A-D**, and discrepancy in boundary calls resulted from the precise location of the boundary being called at one end or the other of the long boundary region. (**L-M**) Effects of MNase digestion level on Micro-C data. Micro-C was carried out at standard MNase levels (1X), and using 2-fold or 4-fold lower levels of MNase, as indicated. Nucleosome laddering for these digestion levels is shown in (L), with substantial di-, tri-, and tetra-nucleosomal DNA being observed in underdigested samples using 0.25X levels of MNase. Marker lanes are 100 bp ladders. Note that the two experiments with 0.5X MNase differ in the

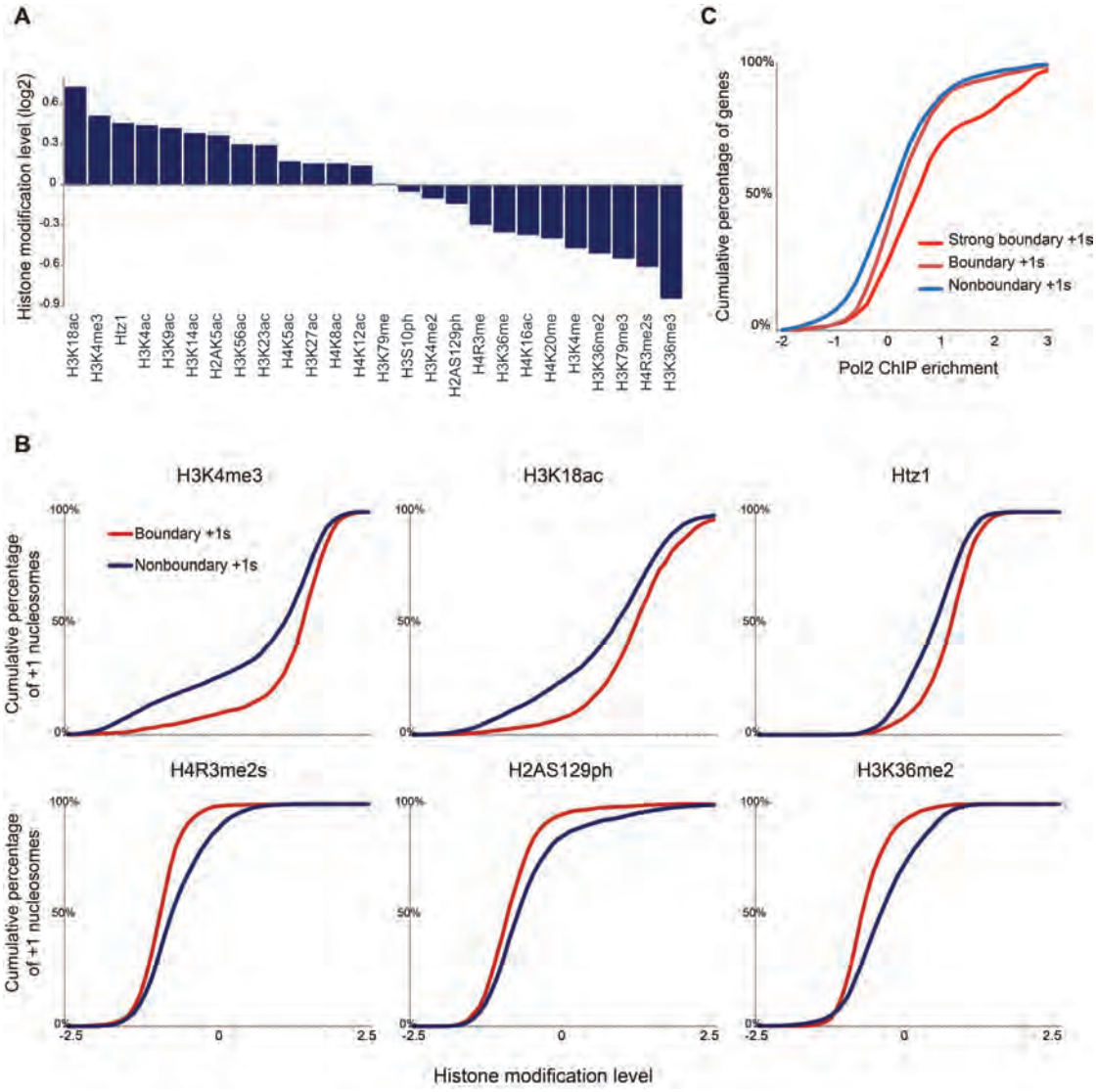
extent of digestion, with replicate 1 exhibiting detectable trinucleosomal DNA. In (M), three plots show Micro-C interaction density vs. distance (each axis in $\log_{10}$ scale) for IN-IN, OUT-OUT, and UNI read orientations, as indicated. Interaction density vs. distance were indistinguishable for all four MNase digestion levels for IN-IN, OUT-OUT, and UNI read pair, while underdigestion resulted in modest changes in abundance of undigested mono and dinucleosomes as seen in the IN-IN plot. Note that the plots of OUT-OUT and UNI are normalized to the reads out to 10 kb as shown in x-axis, while the IN-IN plot is normalized to the reads at the range from 200 bp – 10 kb for eliminating overrepresentation of undigested read pairs. (**N**) Scatterplots showing the number of Micro-C reads crossing an NDR, for 6500 NDRs, for a typical Micro-C dataset (x axis) compared to underdigested Micro-C datasets (y axis). NDR-crossing reads were restricted to read pairs in the IN-OUT or OUT-OUT orientation, with reads being at least 500 bp apart. NDR boundary strength was highly correlated across different digestion levels. These data strongly argue that Micro-C boundaries are not artifacts of overdigested chromatin, although in principle it remains possible that even in the 0.25X dataset digestion of exceptionally “fragile” nucleosomes alters measurable interactions across promoters. (**O**) Single-locus validation of boundaries. For each of the three loci, we designed primers in all four orientations (IN-IN/IN-OUT/OUT-IN/OUT-OUT) for the indicated nucleosomes – +1/-1 and +2/-2 nucleosome pairs flanked NDRs, while the corresponding +1/+3 and +2/+7 pairs were located within the same gene and did not cross a boundary. Note for each case that far more Micro-C product was generated using CID-internal primer pairs than for NDR-crossing primer pairs, and that this was unaffected by MNase digestion level (0.25X vs. 1X MNase levels).

Biochemical features of Micro-C boundaries

What biochemical aspects of a given nucleosome might play a role in boundary activity? Overall, boundary nucleosomes were enriched for the pairs of nucleosomes flanking nucleosome-depleted regions (NDRs) in yeast, and exhibited significant enrichment of a variety of histone marks found at the 5’ ends

of genes (WEINER *et al.* 2015), including high levels of transcription-related marks such as H3K4me3 and H3K18ac (**Figure 2.3A**), and elevated rates of replication-independent H3 replacement (DION *et al.* 2007). Not only were these enrichments significant relative to all nucleosomes in the genome, but for most histone modifications the enrichments observed were also highly significant when comparing boundary +1 nucleosomes only to other +1 nucleosomes (**Figure 2.3B**). Consistent with this, strong boundaries were generally found upstream of more highly-transcribed genes than nonboundary promoters (**Figure 2.3C**). Other features of strong boundaries included high levels of the RSC ATP-dependent chromatin remodeling complex, and high levels of the cohesin loading factor Scc2 (LOPEZ-SERRA *et al.* 2014) (**Figures 2.2A-D**), and these factors were enriched at boundary NDRs relative to all other NDRs (**Figures 2.3D-E**).

These findings are consistent with previous reports (GHELDOF *et al.* 2006; DIXON *et al.* 2012; NORA *et al.* 2012; LE *et al.* 2013) that highly active genes can act as boundaries between self-associating domains. Importantly, the increased resolution afforded by Micro-C localizes such boundaries specifically to active promoters in yeast, thus implicating a number of promoter-specific factors in chromosome folding.


A
Histone modification level (log2)
0.6
0.3
0
-0.3
-0.6
-0.9
H3K18ac
H3K4me3
Htz1
H3K4ac
H3K9ac
H3K14ac
H2AK5ac
H3K56ac
H3K23ac
H4K5ac
H3K27ac
H4K8ac
H4K12ac
H3K79me
H3S10ph
H3K4me2
H2AS129ph
H4R3me
H3K36me
H4K16ac
H4K20me
H3K4me
H3K36me2
H3K79me3
H4R3me2s
H3K36me3
C
Cumulative percentage of genes
100%
50%
0%
-2
-1
0
1
2
3
Pol2 ChIP erichment
Strong boundary +1s
Boundary +1s
Nonboundary +1s
B
H3K4me3
H3K18ac
Htz1
Boundary +1s
Nonboundary +1s
Cumulative percentage of +1 nucleosomes
H4R3me2s
H2AS129ph
H3K36me2
-2.5
0
2.5
Histone modification level

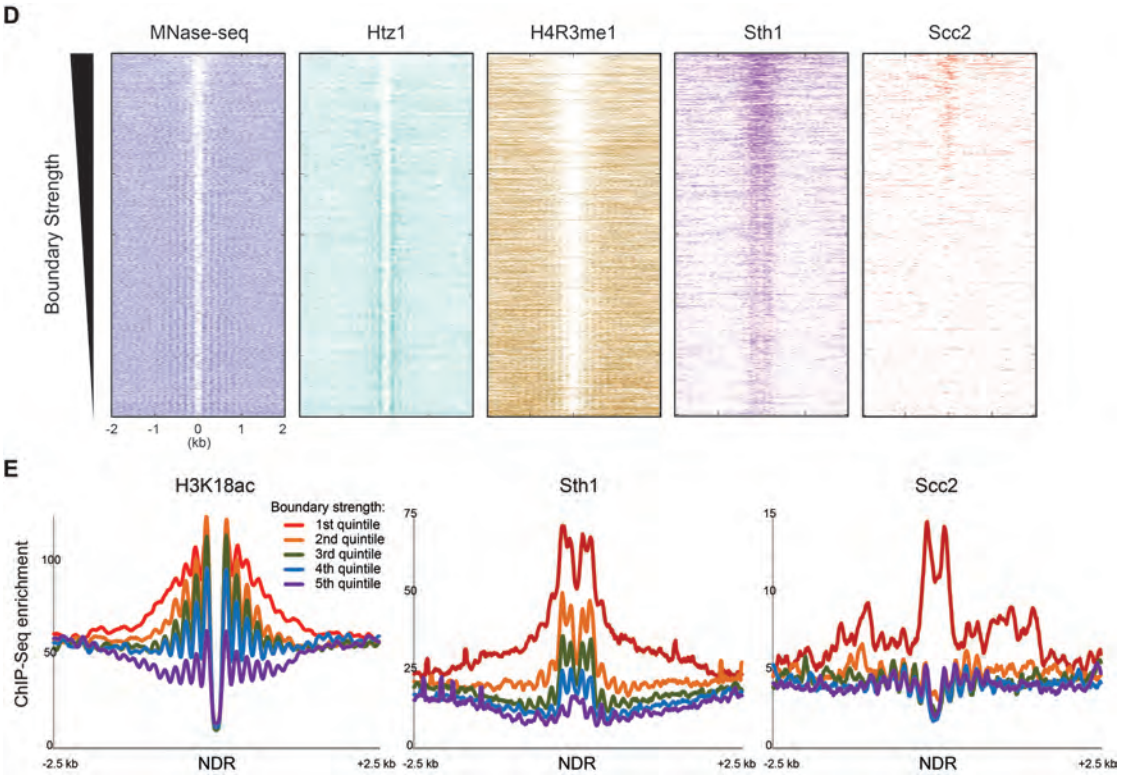


Figure 2.3. Histone modifications and chromatin regulators associated with boundaries.

(**A**) Average histone modification levels are shown for all boundary nucleosomes (using both nucleosomes flanking a boundary). Histone modification data are from (WEINER *et al.* 2015), normalized to nucleosome occupancy, and expressed as $\log_2$ enrichment relative to all nucleosomes genome-wide. (**B**) Boundary +1 nucleosomes are enriched for histone modifications relative to other +1 nucleosomes. Comparison of +1 nucleosomes associated with Micro-C boundaries, vs. all remaining +1 nucleosomes, showing the cumulative distribution for enrichment levels of the indicated histone modifications. In all cases the enrichments are highly ($p < 10^{-10}$) significant by KS test. (**C**) Boundary activity at promoters is associated with elevated transcription rates. For each promoter, the transcription rate is defined as the Pol2 level (KIM *et al.* 2010) of the more highly-transcribed of the two adjacent genes, and cumulative plots show Pol2 enrichment values for nonboundary promoters, for all boundary promoters, and for the strongest half of boundary promoters. (**D**) Heatmaps of chromatin features at boundaries. Each panel shows ChIP-Seq data (or MNase-seq data) for 5 kb surrounding

a Micro-C boundary, with boundaries sorted from strong (top) to weak (bottom). (**E**) Boundary nucleosome-depleted regions (NDRs) are associated with high levels of H3K18ac, RSC, and cohesin relative to all NDRs genome-wide. In each panel, all NDRs in the yeast genome were identified, and Micro-C interactions that cross the NDR were calculated. NDRs were sorted according to quintiles based on the abundance of NDR-spanning Micro-C reads – the first quintile encompasses the 1300 NDRs with the fewest NDR-spanning Micro-C ligation products, etc. – and ChIP-Seq data for the indicated factors was averaged for each quintile.

Properties of gene level of chromatin folding

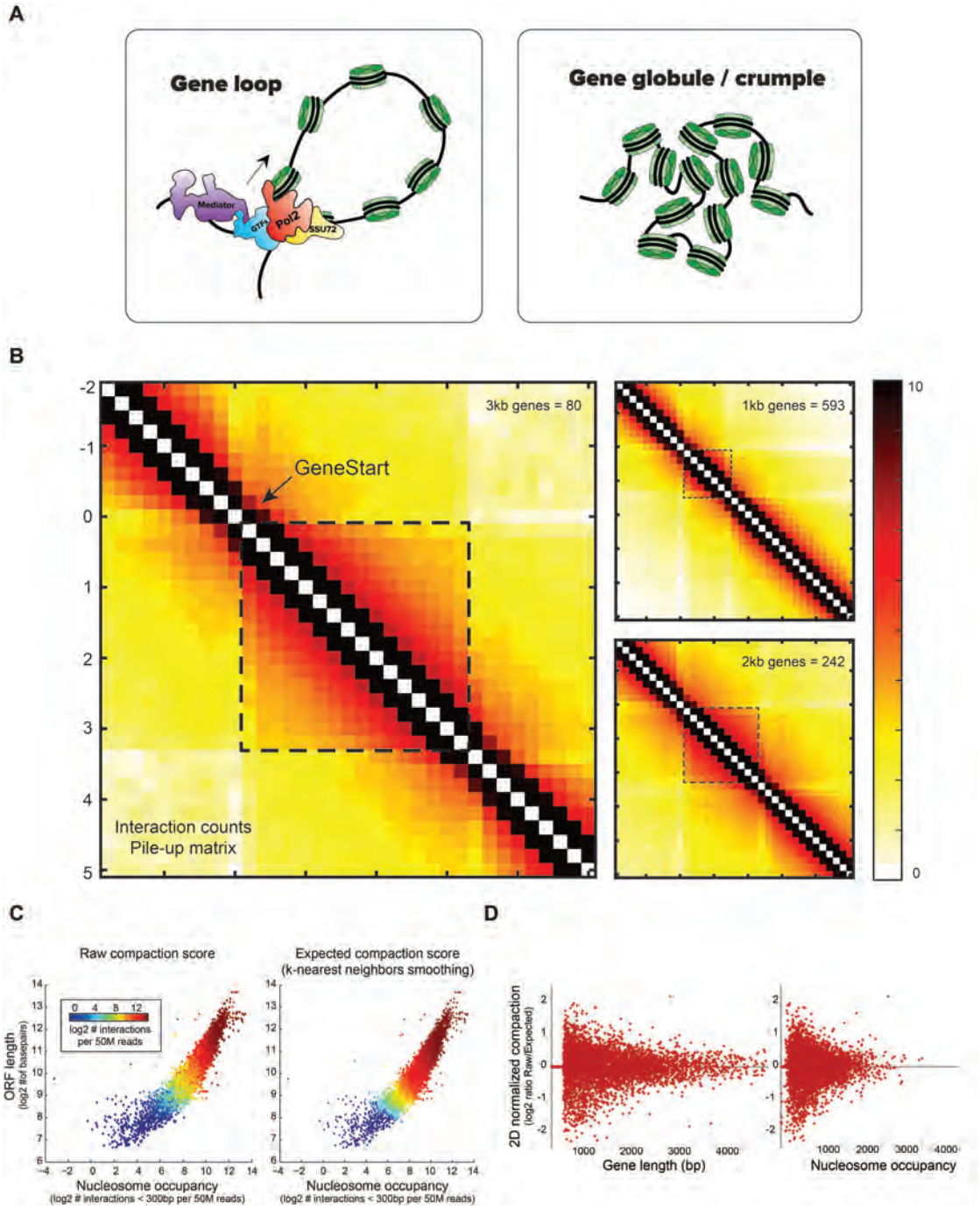
Gene looping structure has been proposed to facilitate transcriptional regulations, e.g. recycle of Pol2 machinery and transcriptional termination, which is mediated by transcription factors, Pol2 CTD phosphatase SSU72, and Mediator complex (**Figure 2.4A**). However, for the vast majority of genes and the genome-wide pile-up analysis we do not observe preferential interaction between a gene's +1 nucleosome and its 3' end nucleosome in Micro-C data, instead finding interactions throughout a gene body (**Figures 2.1B and 2.4B**), suggesting a "gene crumple" or globule rather than a "gene loop" structure (**Figure 2.4A**).

While domains of high local interactions are strongly correlated with gene structure, it is also apparent that genes vary significantly in the extent of such self-association. To identify regulatory and other correlates with gene compaction, we identified genes with significantly high or low levels of self-association (corrected for nucleosome occupancy and gene length – **Figure 2.4C-D**). Overall, gene compaction was anticorrelated (r = -0.56) with transcription rate, even when correcting for nucleosome occupancy (**Figure**

2.4E). The highly-transcribed ribosomal protein genes were associated with particularly open chromatin, not only exhibiting low levels of gene compaction but also commonly forming strong boundaries between CIDs (**Figure 2.2A-D**). Consistent with the anticorrelation between gene compaction and transcription rate, we found that gene compaction was positively correlated with the level of H2AS129ph – a mark of gene repression (SZILARD *et al.* 2010) – associated with the gene, and was anticorrelated with active marks such as H3K4me3, H3K18ac, and others (LIU *et al.* 2005; WEINER *et al.* 2015) (**Figure 2.4I**).

To test the hypothesis that active transcription results in unfolding of genes, we carried out Micro-C in yeast responding to diamide stress, a stimulus which leads to transcriptional changes at ~20% of all yeast genes (GASCH *et al.* 2000). Consistent with the anticorrelation between transcription rate and gene compaction at steady-state, we observed unfolding of genes that were strongly upregulated by diamide stress, and the converse behavior at diamide-repressed genes (**Figures 2.4G-H**). The anticorrelation between transcription rate and chromosome compaction is therefore dynamic, and does not simply reflect disparate regulatory strategies used for transcription of TFIID-dominated “growth” genes and SAGA-dominated “stress” genes during active growth. To further test the hypothesis that transcription leads to unfolding of the chromatin associated with genes, we treated yeast with the RNA polymerase inhibitor thiolutin, finding that highly-transcribed genes in untreated cells – primarily ribosomal protein

genes – became significantly ($p < 6.4 \times 10^{-55}$, t-test) more compact upon inhibition of RNA polymerase (**Figure 2.4F**).



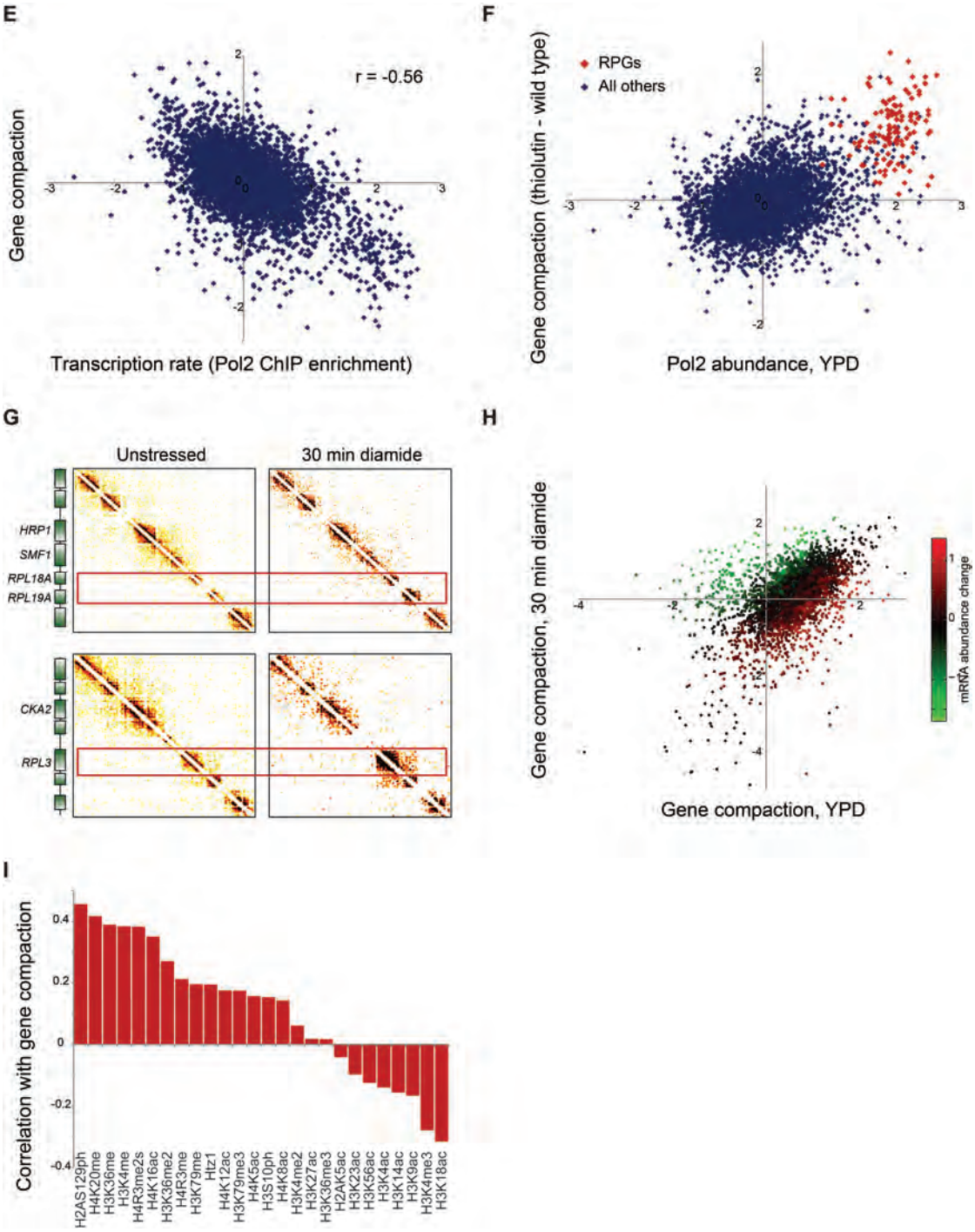


Figure 2.4. Gene-specific compaction is anticorrelated with transcription. (**A**) Schematic models of gene looping and gene globule/crumple. (**Left panel**) Gene looping structure is proposed mediated by transcription machinery including Pol2,

general transcription factors (GTFs), Mediator, and SSU72. (Right panel and **B**) The Micro-C data suggests gene globule/crumple as a common structure in budding yeast. (**B**) Pile-up maps was binned to 200 bp resolution, spanned from 2 kb upstream to 5 kb downstream, and normalized only for sequencing depth. All genes of length 1 – 1.1kb (Top-Right panel), 2 – 2.1 kb (Bottom-Right panel), and 3 – 3.1kb (Left panel), as indicated, were identified and aligned by their 5' ends. The narrow range of gene lengths was chosen to assist in visualization of a discrete 3' gene end in these plots. Gene crumple structure is evident in these plots as a region of increased contacts bounded at both the 5' end and the 3' ends of genes. Note that interactions within each box decay smoothly with increasing distance from the diagonal, indicating that interactions between gene ends are at most a minority subpopulation of gene folding conformations. (**C**) (Left panel) Raw compaction counts per gene (number of interactions within the gene) shown as a function of nucleosome occupancy (x-axis) and gene length (y-axis). Each gene is a point colored according to the raw compaction score. Here, nucleosome occupancy is calculated as Micro-C reads of <300 bp, to internally control Micro-C datasets for both MNase digestion and mutant effects on nucleosome occupancy. (Right panel) Same plot as (Left panel), with compaction score smoothed by neighboring points. This provides the expected compactness for a gene given its length and nucleosome occupancy. (**D**) The normalized compaction score is calculated by taking the ratio of raw score (**C,** Left panel) to the expected (**C,** Right panel) for each gene, and is uncorrelated with (**D,** Left panel) gene length and (**D,** Right panel) nucleosome occupancy. (**E**) Scatterplot comparison of gene compaction with transcription rate. X axis shows Pol2 ChIP data from (KIM *et al.* 2010), y axis shows gene compaction from this dataset based specifically on interactions beyond N/N+3, normalized for nucleosome occupancy and gene length. (**F**) Global inhibition of transcription leads to increased compaction over normally highly-transcribed genes. Here, Pol2 abundance at t=0 is plotted (x axis) against the change in gene compaction in response to 45 minutes of treatment with the RNA polymerase inhibitor thiolutin. Red points show RPGs. (**G**) Changes in transcription affect gene compaction. Micro-C was carried out for yeast subject to 30 minutes of 1.5 mM diamide, a sulfhydryl-reducing agent that alters transcription of ~20% of the yeast genome (GASCH *et al.* 2000). Here, Micro-C contact matrices for unstressed and diamide-stressed yeast show regions surrounding three ribosomal protein-encoding

genes (RPGs), which are strongly repressed in response to diamide stress and which exhibit a dramatic increase in local compaction. (**H**) Systematic analysis of diamide-induced changes in chromosome folding. Here, gene compaction is scatterplotted for unstressed and diamide-stressed yeast, with points color-coded according to the corresponding mRNA abundance changes in diamide (smoothed by 20 nearest neighbors). (**I**) Histone modifications that correlate with gene compaction. Correlation coefficient between normalized gene compaction and the indicated histone modifications (WEINER *et al.* 2015) is plotted on the y axis, with modifications sorted according to their correlation with gene compaction.

Comparison of chromosome folding in various genetic backgrounds

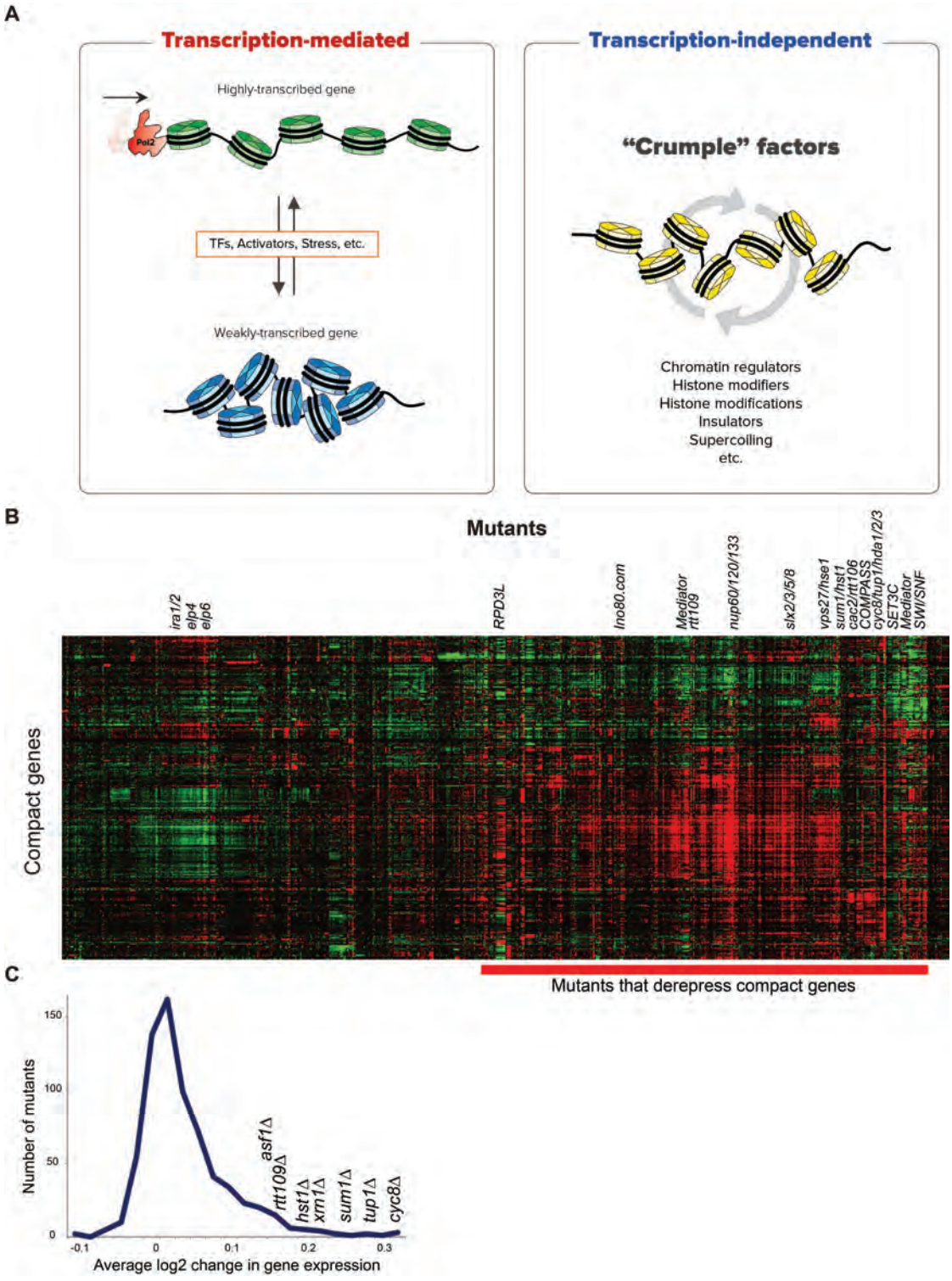
Although transcription can clearly influence gene compaction, transcription rate only explains 31% of the variance in the Micro-C compaction signal, implying that the major influence over gene compaction is related to regulatory strategies or is gene-specific (**Figure 2.5A**). We therefore asked whether compact genes shared any common biology. Comparison of gene compaction with data from 700 yeast deletion mutants (KEMMEREN *et al.* 2014) revealed that unusually compact genes were derepressed in mutants lacking various histone deacetylases (including Cyc8/Tup1 and Sum1/Hst1), histone turnover machinery (such as Rtt109 and Asf1), or Mediator activity (**Figures 2.5B-C**). Based on these results, we carried out Micro-C analysis of several mutants predicted to have strong effects on gene compaction, as well as mutants with plausible roles for gene compaction in the literature but which were not identified based on our analysis of wild-type gene compaction. We screened 24 mutants (including several histone mutants – see below) by Micro-C at relatively low sequencing depth, and chose 14 mutants to

sequence deeper based either on widespread effects on chromosome compaction or on prior findings in the literature.

To visualize mutant effects on gene compaction on a gene-by-gene basis, for each gene we calculated the difference in normalized compaction score between a given mutant and the relevant wild-type. Clustering of this matrix (**Figure 2.5D**, top) revealed that a dominant signal in the first cluster (red) reflects increased compaction of normally highly-transcribed genes such as the ribosomal protein genes (RPGs) upon their repression due to Pol2 inhibition (thiolutin), oxidative stress (diamide), or conditional inactivation of the essential RSC-SCC chromatin remodeler (*sth1-2*, *rsc8-21*, *scc2-4*). In contrast to the effects of transcriptional repression resulting in increased RPG compaction, mutations in another two clusters (blue and green) resulted in mis-compaction of a variety of genes (**Figure 2.5D**, bottom). Importantly, although compaction of RPGs was observed in multiple conditions where these genes are repressed, most mutant effects on gene compaction could not be explained by transcriptional effects of the mutation, as for example the correlations between *rtt109*Δ, *med1*Δ, and *rpd3*Δ effects on gene compaction and mRNA abundance (KEMMEREN *et al.* 2014) were 0.17, 0.03, and -0.006, respectively.

The genome-wide histogram analysis shows that several mutants exhibite widespread changes in gene compaction throughout the genome (**Figure 2.5E-F**). Most notably, we find that genes become less compact, on average, in mutants affecting the Middle/Scaffold modules of Mediator complex, as well as in

mutants lacking the H3K56 acetyltransferase Rtt109 (**Figures 2.5D-F** and **2.5G**). In contrast, chromatin was globally more compact in temperature-sensitive mutants affecting the RSC complex and the cohesin loading factor Scc2 (**Figures 2.5D-F** and **2.5H**). In addition to these major regulators in chromosome folding, loss of many other chromatin regulators, such as the primary RNase H in yeast (Rnh201), lead to more subtle gene-specific defects in gene compaction (**Figures 2.5D-F** and **2.5I**). We next extended our analysis and investigated how these complexes regulate the chromatin folding at single gene level.


A
Transcription-mediated
Highly-transcribed gene
Pol2
TFs, Activators, Stress, etc.
Weakly-transcribed gene
Transcription-independent
"Crumple" factors
Chromatin regulators
Histone modifiers
Histone modifications
Insulators
Supercoiling
etc.
B
Mutants
ira1/2
elp4
elp6
RPD3L
Ino80.com
Mediator
rtt109
nup60/120/133
six2/3/5/8
vps27/hse1
sum1/hst1
cac2/rtt106
COMPASS
cyc8/tup1/hda1/2/3
SET3C
Mediator
SWI/SNF
Compact genes
Mutants that derepress compact genes
C
Number of mutants
150
100
50
0
asf1Δ
rtt109Δ
hst1Δ
xm1Δ
sum1Δ
tup1Δ
cyc8Δ
-0.1
0
0.1
0.2
0.3
Average log2 change in gene expression

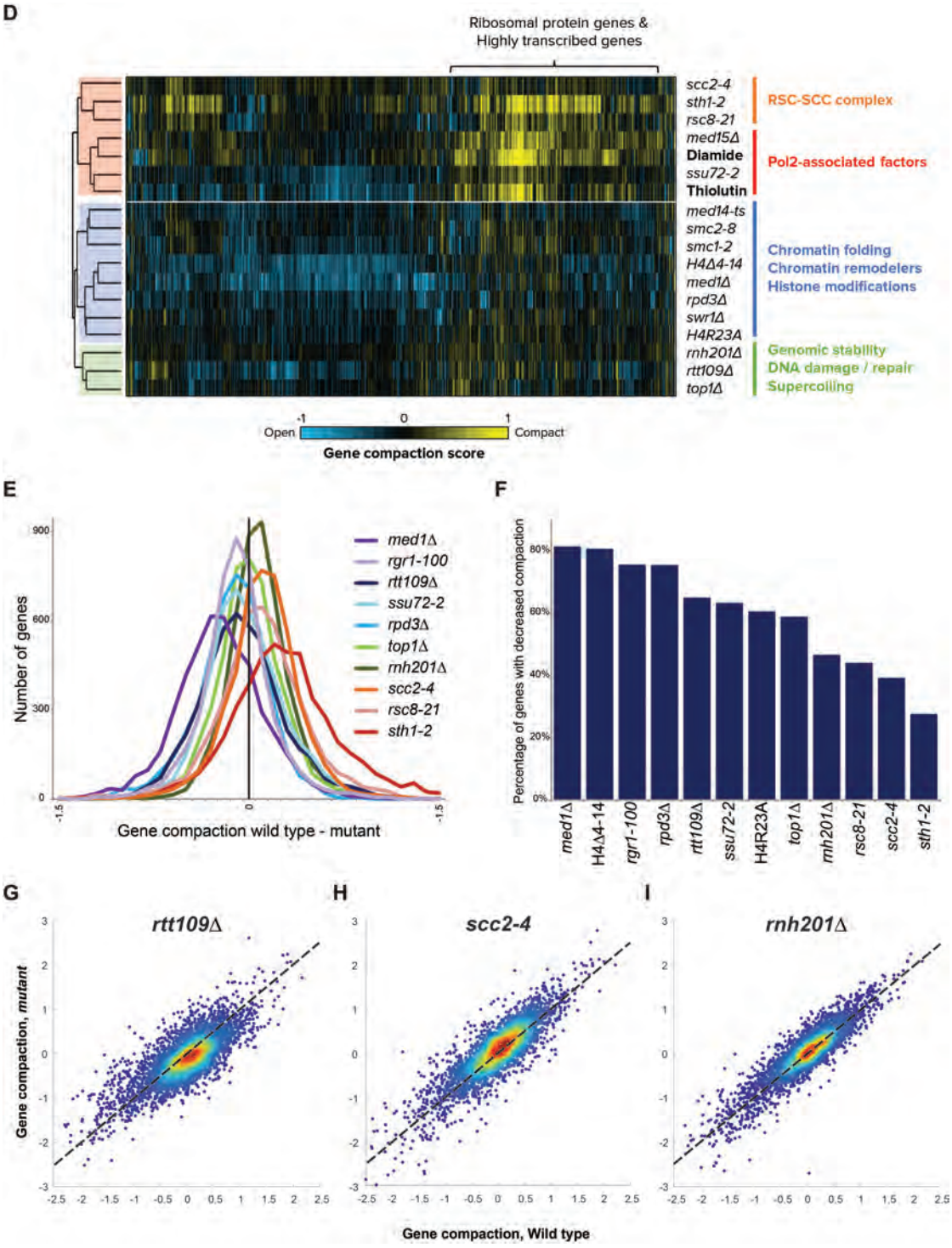


Figure 2.5. Mutant effects on gene compaction.

(**A**) Schematics of transcription-dependent and -independent gene compaction. (Left panel) As shown in **Figures 2.4E-I**, highly-transcribed genes lose gene compaction; in contrast, chromatin gains more compact upon transcriptional down-regulation. (Right panel) In addition to transcriptional regulation, we hypothesized that gene compaction can be regulated by chromatin regulators in a transcription-independent manner. We included chromatin remodelers, histone modifiers, etc. as potential targets for further investigation (see below). (**B**) Heatmap of mutant effects on expression of compact genes. Rows show all 385 highly compact genes, columns show all 700 deletion mutants analyzed. (**C**) Mutants that affect compact genes. For the 385 most compact genes as assayed by Micro-C, the average gene expression change was calculated from data for 700 yeast deletion mutants (KEMMEREN *et al.* 2014). Here, x axis shows the average change in expression for compact genes, with histogram showing the number of mutants (y axis) with the indicated average gene expression change. Right skew of the distribution indicates that more mutants derepress compact genes than induce these genes, due to the fact that highly compact genes tend to be poorly transcribed in wild type yeast (**Figure 2.4E**). Locations of several mutants of interest are indicated along the histogram. (**D**) Global and gene-specific effects of chromatin mutants on gene compaction. For all mutants analyzed by Micro-C, gene compaction scores were calculated, and for all genes with at least a 2-fold change in compaction in one mutant, the difference between all mutants and the relevant wild-type is shown in a clustered heatmap. Note that three major clusters appear in the heatmap as indicated in different colors. The red cluster is dominated by the signal of transcription regulation, while gene compaction in the blue and green clusters are regulated in other strategies. (**E**) Histograms showing the distribution of changes in gene compaction for the indicated mutants or conditions. In each case, nucleosome occupancy-corrected gene compaction was calculated for every gene in the genome, and the difference between BY4741 and the indicated mutants is plotted on the x axis – negative values indicate decreased gene compaction in the mutant, positive values indicate increased gene compaction. (**F**) Mutant effects on compaction. Here, normalized compaction scores were calculated for all genes, and compaction score for each mutant was subtracted from its compaction score in wild type. The number of mutants that “open up” in a given mutant (e.g. with a change in compaction below zero) is shown in a bar graph, with mutants ordered

according to this value. (**G-I**) Scatterplots of occupancy-corrected gene compaction scores for wild type (x axis), and the various indicated mutants. Notable here are a more variable signal with a modest global loss of compaction for *rtt109*Δ (G), a modest global gain of compaction for *scc2-4* (H), and relatively tight scatterplot for *rnh201*Δ (I).

The regulatory mechanisms of chromatin folding

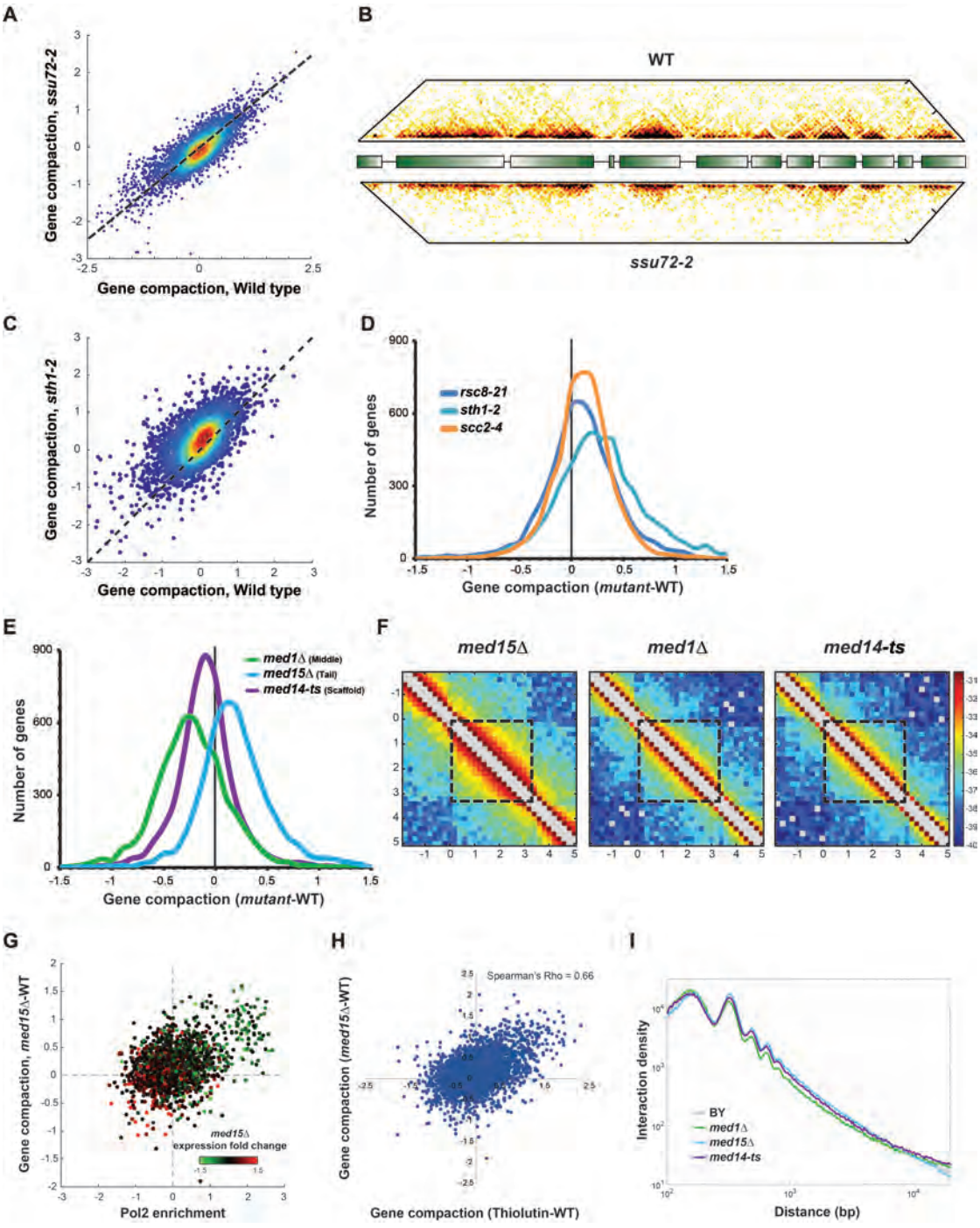
We first investigated the role of the "gene looping" factor Ssu72 (also Pol2 CTD phosphatase) on gene compaction, as the occurrence of domains of strongly self-associating nucleosomes over gene bodies is reminiscent of the gene loops in yeast reported by several groups (O'SULLIVAN *et al.* 2004; ANSARI AND HAMPSEY 2005; TAN-WONG *et al.* 2012). Moreover, gene compaction as measured by Micro-C is anti-correlated with transcription rate (**Figure 2.4E**), in contrast to reported gene loops. Nonetheless, to test the hypothesis that the CID structures observed using Micro-C might be somehow related to gene loops (with technical differences in the protocols revealing different views of the same structure), we performed Micro-C analysis of the *ssu72-2* mutant that lacks detectable gene looping (ANSARI AND HAMPSEY 2005; TAN-WONG *et al.* 2012). We find a moderate but significant decrease in global chromosome compaction, with a corresponding decrease in individual gene compaction, in the *ssu72-2* mutant (**Figures 2.6A-B**).

Chromosome folding effects resulting from inactivation of the cohesin-loader Scc2 were correlated with those resulting from RSC inactivation, consistent with the strong correlations recently reported between the synergetic effects of these mutations on transcription activation and on nucleosome

positioning (LOPEZ-SERRA *et al.* 2014). The RSC and SCC complexes are reported traveling with Pol2 transcription machinery and facilitate elongation as well. Thus the effects could mainly associate with defects of Pol2 transcription (SPAIN *et al.* 2014) (**Figures 2.6C-D**).

Mediator complex encompasses over 25 subunits in budding yeast (TSAI *et al.* 2014; PLASCHKA *et al.* 2015). The uncoupled gene expression profile of the mutants lacking Mediator (VAN DE PEPPEL *et al.* 2005; KEMMEREN *et al.* 2014) and the distinct chromatin binding signatures of each Mediator subunits (FAN *et al.* 2006; GRUNBERG *et al.* 2016; JERONIMO *et al.* 2016; PETRENKO *et al.* 2016) strongly suggest a multitalented role of Mediator in transcription and chromatin remodeling (KAGEY *et al.* 2010; ZHU *et al.* 2011; LIU AND MYERS 2012; NOCK *et al.* 2012; MUKUNDAN AND ANSARI 2013; PHILLIPS-CREMINS *et al.* 2013). Interestingly, we noted that the gene compaction changes in the Mediator mutants deviate from each other – chromatin tends to condense in the Tail mutation (*med15*Δ), while gene crumple is globally disrupted in the Middle mutation (*med1*Δ), and the effects of the Scaffold are in between (**Figures 2.6E-F**). The chromatin compaction of highly- and poorly-expressing genes are more responsive to the transcription changes upon *med15*Δ (**Figure 2.6G**). The positive correlation of gene compaction change between *med15*Δ and Pol2 inhibition (**Figure 2.6H**) further supports the findings that the recruitment of Tail subunit of Mediator to upstream activated regions (UAS) is required for transcription activation. In contrast, the Middle and Scaffold subunits may contribute to general structural

role in chromatin folding at a gene level, but no significant effect on global internucleosome interaction (**Figure 2.6I**). More details will be subjected to future study.



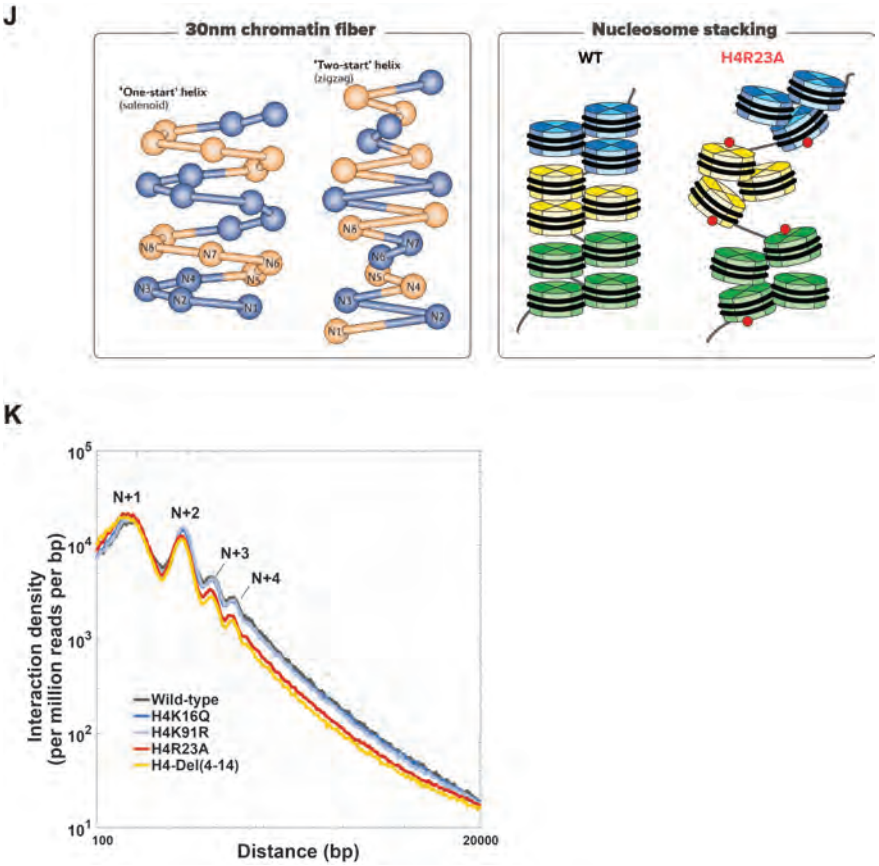


Figure 2.6. The cases of gene looping factor – SSU72, chromatin remodeler/insulator – RSC-SCC2 complex, Mediator complex, and 30 nm chromatin structure.

(**A-B**) Modest effects of ssu72-2 on global gene compaction. In (A), Micro-C was carried out for three replicate cultures of *ssu72-2*. Compaction scores for all genes are scatterplotted for wild-type and *ssu72-2* mutants, as indicated. Gene compaction scores for wild-type and *ssu72-2* mutants shows a subtle but significant ($p < 1.2e-8$, KS test) general loss of gene compaction in the mutant. In (B), two rotated matrixes surrounding ChrIX: 260 kb – 280 kb showing wild type and *ssu72-2* Micro-C data, as indicated. The maps were normalized in the same way as **Figure 2.1B**. (**C-D**) The effects of RSC-SCC2 mutants on gene compaction are transcription dependent. In (C), compaction scores for all genes are scatterplotted for wild-type and *sth1-2* mutants, as indicated. The *sth1-2* mutant globally gains more compaction, particularly in normally highly-transcribed genes. In (D), histograms of mutant effects on gene compaction, as in

Figure 2.5E, showing that all three mutants lacking the activity of RSC-SCC2 complex increase gene compaction, consistent to their synergetic effects on transcription. (**E-I**) Distinct effects of the Mediator subunits on chromatin folding. In (E), histograms of mutant effects on gene compaction, as in **Figure 2.5E**. Mutants in the middle, tail, and scaffold subunits of Mediator complex, as indicated. Notable here are a global gain of compaction for the tail mutant (*med15*Δ), a drastic loss of compaction for the middle mutant (*med1*Δ), and a modest change for the scaffold mutant (*med14-ts*). In (F), pile-up matrixes were drawn as in **Figure 2.4B** but in blue-to-red color scheme for clear visual comparison, indicating higher level of gene compaction in *med15*Δ than in *med1*Δ and *med14-ts*. In (G), Pol2 abundance (x axis) is plotted against the change in gene compaction in *med15*Δ (y axis) and overlaid an additional green-to-red heatmap shown as the change in $\log_2$ gene expression in *med15*Δ. Here shown the gene compaction change in the tail mutant is highly associated with the transcriptional regulations. In (H), Change in compaction score for all genes are scatterplotted for the treatment of polymerases inhibitor Thiolutin (x axis) and *med15*Δ (y axis). A positive correlation (spearman rho = 0.66) in the gene compaction change argues that the tail subunit of Mediator regulates chromatin folding in a transcription-dependent manner. In (I), Interaction density vs. distance decay curve was calculated as in **Figure 2.1C**, shown no significant change in internucleosomal interactions in *med15*Δ and *med14-ts*Δ, but a moderately decrease in *med1*Δ. (**J-K**) Micro-C data support the model of tetra-nucleosome stacking, instead of periodic 30 nm chromatin structure. In (J), (Left panel) schematic of two hypothesized paths of 30 nm chromatin structure. The one-start (solenoid) helix is formed by average six nucleosomes per helical turn, while the two-start (zig-zag) helix turns periodically in every three nucleosomes. (Right panel) Instead of periodic 30 nm chromatin fiber, one model proposed tri- or tetra-nucleosomal stacking mediated by histone H4 tail and H4R23 methylation. In (K), interaction density vs. distance for yeast strains carrying either a single copy of H3/H4 (DAI *et al.* 2008), or carrying the indicated histone mutants, shown as in **Figure 2.1C**. Shown here are “UNI” read pairs only, with data normalized to all reads in distance up to 20 kb. H4K16Q, H4R23A, and H4Δ4-14 have all been implicated in chromosome folding in vitro, while H4Q93E was included as an unrelated control.

Systematic analysis of short-range nucleosomal interactions

Finally, we turn to short range aspects of chromosome folding. The dominant models for the 30 nm fiber are the “zig-zag” (DORIGO *et al.* 2004; SONG *et al.* 2014) and solenoid models (GHIRLANDO AND FELSENFELD 2008), which differ in their periodicity – zig-zag models predict that nucleosomes N and N+2 (N+4, N+6, etc.) should be in spatial proximity, while interdigitated solenoid models typically have a periodicity of ~5-6 nucleosomes. We do find some support for the possibility that a common motif of individual tri- or tetra-nucleosomal zig-zags (SCHALCH *et al.* 2005; SONG *et al.* 2014) may exist in vivo, as N/N+2 nucleosome pairs are nearly as abundant as N/N+1 nucleosome pairs genome-wide, particularly when excluding ligation products with “in-in” read pairs (**Figure 2.1C** and **2.1G**). However, plotting the decay of nucleosome-nucleosome interactions as a function of distance in our Micro-C dataset reveals no evidence for long-range periodicity in internucleosomal interactions (**Figure 2.1C** and **2.1G**). Our data therefore do not support a periodic repeating fiber, but are consistent with the idea of a tri or tetranucleosomal motif in chromatin fiber folding – in this model, the absence of high levels of N/N+4 and N/N+6 in **Figure 2.1C** could reflect either lack of extended zig zag stretches in vivo (eg a given stretch of 12 nucleosomes might only carry a single folded tetranucleosome in a given cell), or, more likely, a technical inability to connect nucleosomes at longer distances due to the use of the short-distance crosslinker formaldehyde (GRIGORYEV *et al.* 2009) (**Discussion**).

To further explore short-range structure in chromosome folding, we assayed chromatin folding in several histone mutants that alter internucleosomal interactions in vitro, as well as in unrelated histone mutants as controls. Most notably, as both H4K16 acetylation and histone N-terminal tails have been implicated in 30 nm fiber formation (SHOGREN-KNAAK *et al.* 2006; HIZUME *et al.* 2009), we carried out Micro-C in relevant mutants to dissect their roles in chromosome folding. As expected, we find that partial deletion of the H4 N-terminal tail leads to a strong loss of chromosome folding signal (**Figures 2.6J**), consistent with a key role for histone tails in chromosome compaction. In contrast, we find no global effects of the H4K16Q mutation on overall chromosome folding, although this is perhaps not surprising given that the majority of the budding yeast genome is euchromatic and is packaged into H4K16-acetylated histones. Finally, we also examined chromosome folding in H4R23A mutants, which do not influence tetranucleosome folding in vitro but alter stacking of adjacent tetranucleosomes (SONG *et al.* 2014). Intriguingly, these mutants exhibited normal interactions up to the tetranucleosome scale but lost interactions beyond this scale (**Figures 2.6J-K**), the length scale at which tetranucleosome packing effects would manifest themselves.

DISCUSSION

We describe here a method for nucleosome-resolution chromosome folding studies, providing genome-wide access to the length scale between 100 bp and

~10kb. We do not find evidence for a repeating 30 nm fiber structure in vivo, although the pattern of short-range interactions we observe is consistent with a tri- or tetranucleosome folding motif in eukaryotic chromatin. Instead, the primary level of organization in the yeast genome observed here is associated with gene structure, with domains of 1-5 genes forming compact gene crumples, or globules, rather than loops (**Figure 2.7**).

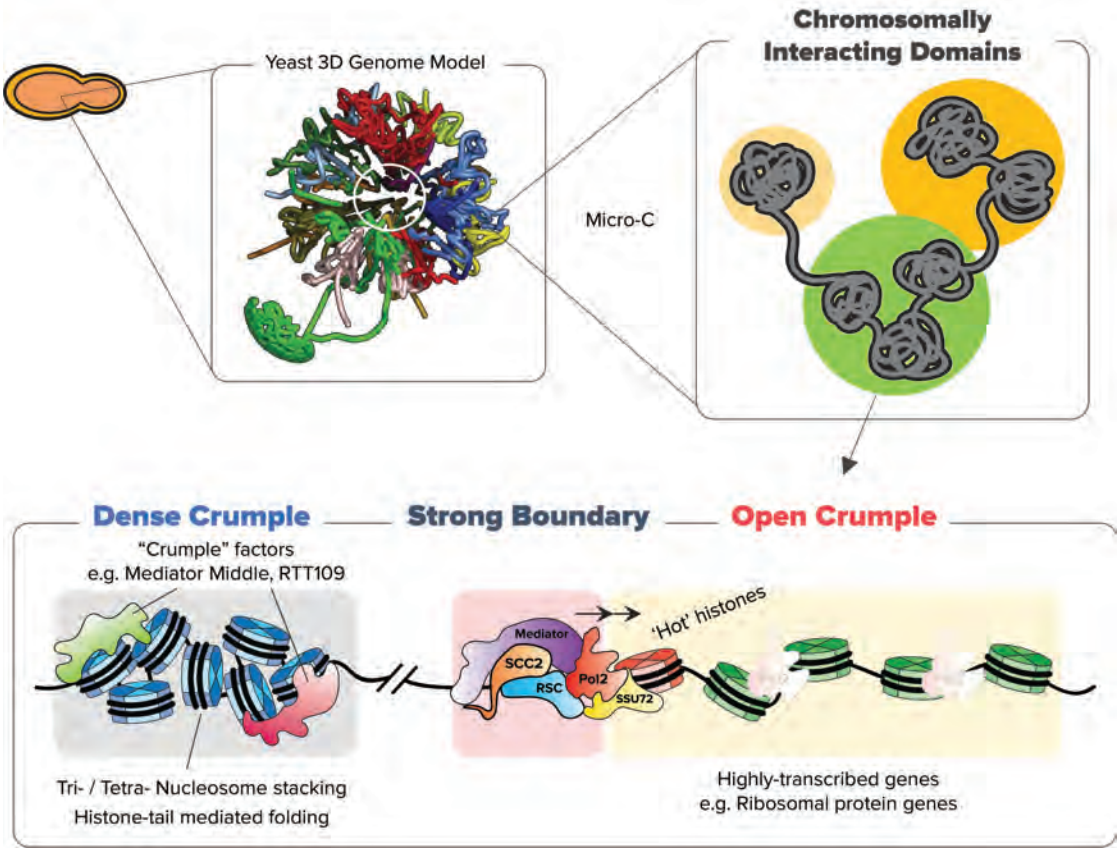


Figure 2.7. Nucleosome-resolution view of chromosome folding.
Cartoon in the upper panel showing global yeast chromosomes are organized as Rabl-like structure - which centromeres are clustered together as a focal point at nuclear periphery and telomeres are spread as multiple loci surrounding nuclear periphery. In this study, the fine-resolution of Micro-C afforded to identify CIDs/TADs in yeast, which generally encompasses 2 – 10 kb of DNA, 1 – 5 genes. The bottom panel illustrating a

model how chromatin organized at the single-nucleosome level. Active transcription including Pol2 machinery, RSC-SCC2 complex, and active histone marks create a phase facilitating boundary forming. Gene level of folding is mainly dominated by transcription activity as highly-transcribed gene with loose compaction and poorly-transcribed genes with dense compaction. “Crumple factor” may contribute to additional layers to regulate gene folding in a transcription-independent strategy. More important, nucleosomes tend to form H4 tail-mediated tri- or tetra-nucleosomal stacking, instead of periodic 30 nm chromatin fibers.

Chromosomal interaction domains in yeast

Our data demonstrates that the self-associating domains observed in many other species are also present in budding yeast, and that this hierarchical folding holds even at the previously unobservable ~2-10 nucleosome scale. The relatively short length of yeast genes presumably explains the prior inability to observe CIDs in this organism (DUAN *et al.* 2010), and suggests that Micro-C might reveal CID structures in organisms such as *Arabidopsis* where they have not been previously observed (FENG *et al.* 2014). Perhaps the most surprising aspect of our study is the finding that the typical length of CIDs tends to be conserved between species when scaled by gene number, rather than linear distance – yeast CIDs typically cover ~1-5 genes, or ~5 kb, while mammalian TADs also cover roughly the same number of genes, thus encompassing ~0.5 Mb. This disparity in length scale suggests that the primary determinant of chromosome folding may be the boundaries that separate compacted domains, rather than the internal structure of the domains themselves, as the detailed folding of the “beads on a string” within a given domain likely varies quite a bit between

different organisms. In other words, we speculate that establishment of boundaries that separate chromosomal domains is the driver of chromosome folding behavior, with folding of the chromosome within each boundary-delimited domain not necessarily conforming to any regular secondary structure.

Prior Hi-C studies have localized boundaries between folding domains to several features of the underlying sequence, including binding sites for the insulator CTCF (DIXON *et al.* 2012; PHILLIPS-CREMINS *et al.* 2013), highly transcribed coding regions (DIXON *et al.* 2012; NORA *et al.* 2012; SEXTON *et al.* 2012; LE *et al.* 2013), binding sites for cohesin (PHILLIPS-CREMINS *et al.* 2013; MIZUGUCHI *et al.* 2014), and sites of Mediator localization (PHILLIPS-CREMINS *et al.* 2013). In addition, insulator function in flies has been linked to regions of dynamic chromatin (NEGRE *et al.* 2010). Although *S. cerevisiae* does not appear to encode a CTCF-like regulator, other features of boundaries between compacted domains appear to be conserved, as boundaries in our dataset are enriched for active promoters and for regions associated with the RSC and cohesin loading complexes (**Figures 2.2-2.3**). Moreover, the increased resolution afforded by Micro-C allows us to localize the boundary activity of highly-transcribed genes specifically to their promoters.

Mutant effects on chromosome compaction

To uncover functional roles for chromatin regulators in establishing chromatin domains in yeast, we analyzed a number of deletion and temperature-sensitive

genes using Micro-C. In general, we noted that many of the features enriched at CID boundaries play functional roles in gene compaction, as we observe a general loosening of chromosome structure in *med1*Δ and *rtt109*Δ mutants. Conversely, *sth1*^{ts} and *scc2*^{ts} mutants shifted to the restrictive temperature exhibited increased gene compaction, presumably due to the similar effects of these mutants on global transcription (LOPEZ-SERRA *et al.* 2014). The key role for Mediator in gene compaction in yeast described here is of great interest, as Mediator recruitment of cohesin has previously been shown to play a role in chromosome folding in murine ES cells (KAGEY *et al.* 2010; PHILLIPS-CREMINS *et al.* 2013), suggesting that chromosome domain compaction may be a conserved consequence of Mediator action. More novel is the finding that Rtt109, a H3K56 acetyltransferase which enhances replication-independent histone turnover (RUFIANGE *et al.* 2007; KAPLAN *et al.* 2008), has global (albeit more subtle) effects on chromosome compaction.

Taken together, our mutant analyses identify a number of factors that help compact genes in budding yeast, and emphasize the key role that promoters play in establishing the folding of the chromosome.

Insights into secondary structure in vivo

A key goal in development of Micro-C was to shed light on chromatin fiber folding, as the dominant models for regular 30 nm fiber structure – zig-zag and solenoid – make distinct predictions for periodicity in internucleosomal contacts.

Here, we find no evidence for periodicity in short-range internucleosome interactions, consistent with the hypothesis that a regular 30 nm fiber may be uncommon in vivo (DEKKER 2008). However, despite the lack of periodicity in our dataset, we do find that N/N+1 and N/N+2 ligation products are present at similar abundance in vivo (**Figure 2.1G**), broadly supporting two-start models for chromatin secondary structure. Indeed, EM studies of folded 12 nucleosome templates subject to limited formaldehyde crosslinking reported similar abundance of N+1 and N+2 crosslinking products, with relatively few interactions at N+4 or greater distances (GRIGORYEV *et al.* 2009). Thus, the absence of the excess of N+4 and N+6 products expected from a regular zig-zag fiber could potentially result from individual tri- or tetra-nucleosome folding motifs (SONG *et al.* 2014) occurring sparsely, and could also result from a technical limitation in Micro-C – formaldehyde is a short distance crosslinker, and chromatin in 30 nm fiber might not present primary amine groups in close enough proximity to be ligated to one another. Suggesting the latter hypothesis, we find that the H4R23A mutant previously shown to affect tetranucleosome stacking, but not the tetranucleosome structure itself (SONG *et al.* 2014), causes a subtle relaxation of chromatin by Micro-C (**Figures 2.6J-K**).

Thus, whether the lack of periodicity in Micro-C data results from a technical inability to capture N/N+4 or N/N+6 interactions from a fully-folded fiber, or whether it reflects the sparse existence of tetranucleosomes in vivo which seldom stack with one another, remains to be tested (potentially via use of

alternative crosslinkers, or examination of higher order multi-nucleosome ligation products). Technical benchmarking of Micro-C using defined 30 nm templates in vitro would be valuable (GRIGORYEV *et al.* 2009), although as the majority of such templates at present use repeats of the Widom601 sequence to ensure uniform nucleosome occupancy, sequencing-based assays cannot distinguish between nucleosome positions along such templates.

With such technical caveats noted, we do note that the lack of periodic Micro-C signal is consistent with the fact that regular 30 nm signatures are seldom observed in EM studies of intact or sectioned nuclei (MCDOWALL *et al.* 1986; NISHINO *et al.* 2012). Moreover, the massive differences in the length scales of CIDs between yeast and mammals are difficult to reconcile with a requirement for a regular form of secondary structure in chromosome folding.

Perspective

Taken together our data support the idea that eukaryotic chromosomes are thus subject to hierarchical “beads of a string” architecture, with the 10 nm nucleosome-linker-nucleosome fiber being the first level of beads on a string, and gene crumples separated by regions of high histone turnover being the next level of organization (**Figure 2.7**). These studies provide a high resolution view of the eukaryotic genome, and, given the powerful tools available in budding yeast, should provide fertile ground for future genetic interrogation of chromosome folding in vivo.

Materials and Methods

Chromatin digestion and end repair. Yeast were grown to midlog phase in YPD media at 30 C, fixed with 3% formaldehyde for 15 min, and quenched with 125 mM glycine for 5 min. Cells were pelleted, spheroplasted with Zymolyase, and MNase-treated to yield >95% mononucleosomes (Liu *et al.* 2005). After stopping MNase, chromatin supernatant was concentrated in an Amicon 10K spin column and dephosphorylated using Antarctic phosphatase. Crosslinked chromatin was subject to T4 DNA polymerase with ATP to leave 5' single-stranded termini, then biotinylated dsDNA was generated by supplementing with biotin-dCTP, biotin-dATP, dTTP and dGTP.

Proximity ligation. 0.5 - 1 µg of crosslinked chromatin was diluted to 10 mL (0.05 – 0.1 µg/mL final) and treated with T4 DNA ligase. After heat inactivation, chromatin was concentrated to 250 µL in an Amicon 30k spin column, and treated with 100U exonuclease III for 5 min to eliminate biotinylated ends of unligated DNA. Proteinase K was then added and incubated for 65 C overnight. DNA was purified by PCI extraction and ethanol precipitation, treated with RNase A, and ~250-350 bp DNA was gel-purified.

Library preparation and sequencing. Purified DNA was treated with End-it, subject to A-tailing with Exo- Klenow, and ligated to Illumina adaptors. Adaptor-

ligated DNA was purified with streptavidin beads to isolate ligated Micro-C products away from undigested dinucleosomal DNA. Streptavidin beads were then subject to ~12-15 cycles of PCR using Illumina paired-end primers. Amplified library was purified and subject to Illumina HiSeq paired end sequencing.

Sequence analysis and mapping. Paired 50 bp reads were extracted and mapped individually to the *S. cerevisiae* genome (sacCer3 genome build) with the bowtie2 tool using the following command line:

```
bowtie2 -x S288C_reference_sequence_20110203 -U <input
fastq> -S output.sam -p 16 -a -–mm –-reorder
```

For a pair of reads mapping to more than one possible location, we selected the shortest interaction. We discarded repeat occurrences of the same fragment pair to avoid PCR artifacts, resulting in unique mappable fragments as the raw interaction data**.** For some visualizations, fragments pairs were associated with one of the 66,360 nucleosomes in budding yeast, while raw fragment pairs were used for analyses such as bp-resolution interaction vs. distance plots (**Figure 2.1C**) and boundary calls (**Figures 2.2-2.3**). In some analyses as indicated, ligation products with IN-IN reads (< 300 bp) were discarded for clear visualizations, as these include an excess of undigested dinucleosomes (**Figure 2.1G**).

Normalization and genome wide interaction analysis. To deal with outliers due to repetitive sequences, we counted the number of interactions in the merged WT samples (using tiled 100bp regions). Regions with more than 10X interactions than the window average were excluded for further analyses. In total 1.3% of the genome was excluded, mostly within Chromosome 12 in the rDNA region, and in Ty elements. Interaction count matrices were normalized to the total number of unique mapped fragments. Correction for nucleosome occupancy, either by iterative row normalization or by normalizing to measured nucleosome occupancy, had minimal effects on Micro-C contact maps (not shown).

Detecting domain boundaries. To detect domain boundaries, we first define a local boundary score for genome position i as the number of interactions of distance *500-10,000* base-pairs that pass above position *i* (i.e. interaction between position *i-d1* and *i+d2*, where *500<d1+d2<10,000)* We then find local minima in this vector of counts as boundaries. The strength of the boundary is inversely proportional to the number of interactions crossing it (**Figures 2.2G-J**).

NDR-centric analysis was carried out for all NDRs in yeast, with the number of NDR-crossing reads (excluding IN-IN reads and read pairs <500 bp apart). For analyses in **Figure 2.2N** and **Figure 2.3**, the 219 NDRs with the most NDR-crossing reads were excluded, as they were primarily associated with rDNA repeats, subtelomeres, and Ty elements.

Compactness score. The raw compaction score, defined as the number of interaction fragments (>300bp) within a gene, is highly correlated with gene size and average nucleosome occupancy (**Figures 2.4C-D**). As a proxy for nucleosome occupancy we use the number of interaction (<300bp) over the gene, as these are mostly due to non-ligated mononucleosome fragments. This score internally normalizes for mutant and MNase digestion effects on nucleosome occupancy. To normalize for these two related measures (gene length and nucleosome occupancy) we used a *k*-nearest-neighbors smoothing with a Gaussian kernel to obtain the expected compactness score for each size and occupancy bin (**Figure 2.4C**). We defined the compactness score of a gene as the log of the ratio between the actual value and the smoothed value. The resulting score is independent of both gene length and occupancy (**Figure 2.4D).**

CHAPTER III

Micro-C XL: Assaying Chromosome Conformation at Length Scales from The Nucleosome to The Entire Genome

Abstract

Structural analysis of chromosome folding in vivo has been revolutionized by Chromosome Conformation Capture (3C) and related methods, which use proximity ligation to identify chromosomal loci in physical contact. We recently described a variant 3C technique, Micro-C, in which chromatin is fragmented to mononucleosomes using micrococcal nuclease, enabling nucleosome-resolution folding maps of the genome. Here, we describe an improved Micro-C protocol using long crosslinkers, termed Micro-C XL, which exhibits greatly increased signal to noise, and provides further insight into the folding of the yeast genome. We also find that signal to noise is much improved in Micro-C XL libraries generated from relatively insoluble chromatin as opposed to soluble material, providing a simple method to physically enrich for bona-fide long-range interactions. Micro-C XL maps of the budding and fission yeast genomes reveal both short-range chromosome fiber features such as chromosomally-interacting domains (CIDs), as well as higher-order features such as clustering of centromeres and telomeres, thereby addressing the primary discrepancy between prior Micro-C data and reported 3C and Hi-C analyses. Interestingly, comparison of chromosome folding maps of *S. cerevisiae* and *S. pombe*

revealed widespread qualitative similarities, yet quantitative differences, between these distantly-related species. Micro-C XL thus provides a single assay suitable for interrogation of chromosome folding at length scales from the nucleosome to the full genome.

Introduction

The compaction and organization of the physical genome has wide-ranging consequences for genomic function (HORN AND PETERSON 2002; DEKKER *et al.* 2013; DEKKER AND MISTELI 2015; FRIEDMAN AND RANDO 2015). In eukaryotes, the first level of genome compaction is organization into the characteristic “beads on a string” structure, with nucleosomes separated by relatively accessible linker DNA. Our understanding of this primary structure of chromatin is well-developed, with multiple crystal structures solved for the nucleosome (LUGER *et al.* 1997; KORNBERG AND LORCH 1999), and a plethora of genome-wide studies that identify the positions of individual nucleosomes across the genome in various organisms, in some cases at single nucleotide-resolution (HUGHES AND RANDO 2014). The next step in chromosome folding remains relatively poorly-characterized; for example, the long-held belief that chromatin fibers form a helical secondary structure termed the 30 nm fiber is increasingly subject to debate (DORIGO *et al.* 2004; TREMETHICK 2007; GHIRLANDO AND FELSENFELD 2008; ROUTH *et al.* 2008; GRIGORYEV *et al.* 2009; COLLEPARDO-GUEVARA AND SCHLICK 2014; SONG *et al.* 2014; CHENG *et al.* 2015).

Structural analysis of chromosome folding beyond the nucleosome fiber has been revolutionized by the Chromosome Conformation Capture (3C) family of techniques (DEKKER *et al.* 2002; DEKKER *et al.* 2013). In 3C-based protocols, chromatin is first crosslinked in vivo using formaldehyde to capture physical interactions between distal regions of the genome. Chromatin is subsequently fragmented, and ligation of chromatin fragments is used to generate chimeric DNA molecules. Sequencing these molecular libraries provides a readout of genomic loci that were crosslinked to one another via protein-protein interactions. Genome-wide variants of 3C, such as Hi-C, have revealed a number of organizational features of eukaryotic genomes at increasingly fine resolutions, from the scale of full chromosomal territories, to multi-Mb active and inactive compartments, to hundred-kb contact domains (TADs), to enhancer-promoter loops (LIEBERMAN-AIDEN *et al.* 2009; DIXON *et al.* 2012; LI *et al.* 2012; NORA *et al.* 2012; SEXTON *et al.* 2012; LE *et al.* 2013; MARBOUTY *et al.* 2014; MIZUGUCHI *et al.* 2014; RAO *et al.* 2014; EAGEN *et al.* 2015; HSIEH *et al.* 2015; WANG *et al.* 2015). While many factors impact the effective resolution of a 3C/Hi-C dataset, including sequencing depth and library complexity (LAJOIE *et al.* 2015), a fundamental limit to genomic resolution is the size of the fragments generated before physical interactions are captured via ligation. Since the majority of 3C-based experiments rely on restriction enzymes for fragmentation of the genome – resulting in genomic fragments that are both long relative to the nucleosome, and

inhomogeneously spaced along the genome – current Hi-C datasets are limited to ~1 kb resolution.

To improve the resolution of 3C-based techniques, we recently developed a high resolution 3C-based technique, dubbed “Micro-C”, in which fragmentation of the genome is accomplished using micrococcal nuclease (MNase) to enable mononucleosome-resolution analysis of chromosome folding (HSIEH *et al.* 2015). While the improved resolution afforded by Micro-C enabled the identification of features such as chromosomally-interacting domains – “CIDs” – in budding yeast that had not previously been discernible using a restriction enzyme-based 3C technique (DUAN *et al.* 2010), known higher-order interactions such as centromere clustering were poorly recovered using our technique. In addition, although several studies have reported “gene loops” in budding yeast using 3C methods (O'SULLIVAN *et al.* 2004; ANSARI AND HAMPSEY 2005), we found no evidence for gene loops using Micro-C.

These discrepancies with the literature motivated a deeper exploration of the effects of specific protocol steps on the results of Micro-C analysis of chromosome folding. Most notably, we sought to determine whether the reliance on formaldehyde, a “zero length” crosslinker, to crosslink genomic loci to one another might limit the ability of 3C-related methods to fully interrogate chromosome structure. To investigate whether longer crosslinkers might reveal additional features of local chromatin structure, we characterized the effects of two long protein-protein crosslinkers on Micro-C maps of the budding yeast

genome. A revised Micro-C protocol incorporating long crosslinkers, which we named “Micro-C XL”, not only recapitulated the local chromatin structures previously revealed by Micro-C, but also robustly recovered higher-order features such as centromere-centromere interactions. Micro-C XL thus overcomes the key technical limitation of the original Micro-C protocol, providing a single protocol for analysis of chromosomal folding from the scale of nucleosomes to the full genome. We also characterized Micro-C XL profiles in pellet and supernatant fractions of crosslinked chromatin, finding that chromatin contacts are enriched in relatively insoluble chromatin, thereby providing a simple technical approach to improve signal-to-noise in Micro-C maps. Finally, we compared Micro-C XL maps from *S. cerevisiae* and *S. pombe*, finding a general conservation of gene-scale folding behavior in these distantly-related species. Taken together, our results provide an updated Micro-C protocol for characterization of chromosome folding at all length scales, and provide additional high resolution insights into chromosome structure in two key model organisms.

Result

Optimization of crosslinking conditions for Micro-C

We recently detailed a modified Hi-C protocol (LIEBERMAN-AIDEN *et al.* 2009) termed Micro-C, in which micrococcal nuclease (MNase) digestion of crosslinked chromatin enables the analysis of chromosome folding at mononucleosomal resolution (HSIEH *et al.* 2015). Our reported Micro-C maps robustly captured

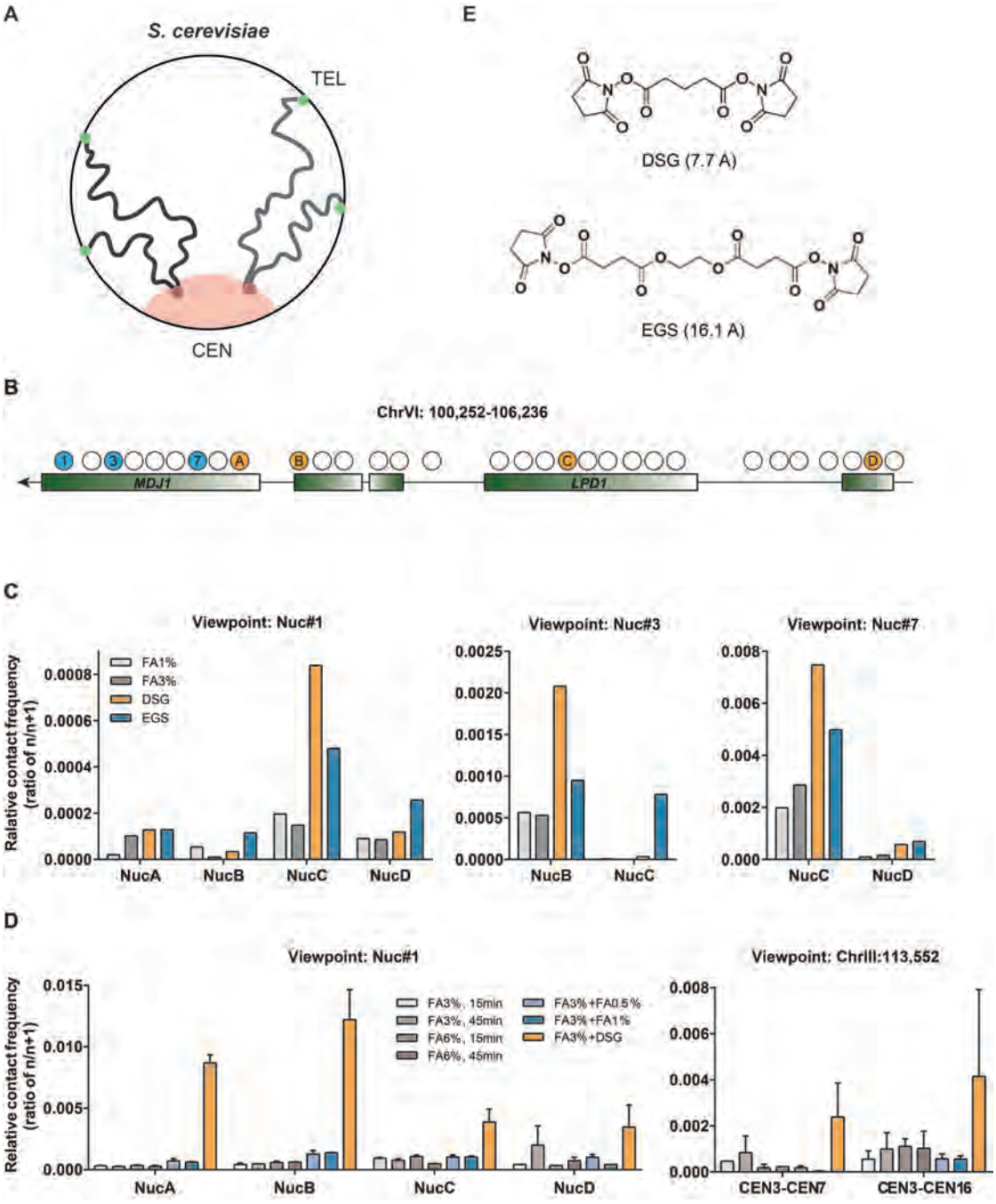
short-range interactions such as chromosomally-interacting domains (CIDs) in budding yeast, but exhibited poor recovery of higher-order features such as the centromere-centromere (CEN-CEN) and telomere-telomere (TEL-TEL) interactions that are well-known features of yeast genome organization (**Figure 3.1A**).

These discrepancies with the literature motivated a deeper exploration of the effects of specific protocol steps on the results of Micro-C analysis of chromosome folding. Most notably, we hypothesized that reliance on formaldehyde, a “zero length” crosslinker, to crosslink genomic loci to one another might limit the ability of 3C-related methods to fully interrogate chromosome structure. Using q-PCR primers designed to assay interactions either within the contact domain associated with *MDJ1* or between pairs of centromeres (**Figure 3.1B-D**), we tested a variety of different protein-protein crosslinkers and crosslinking conditions to identify conditions that best enabled recovery of longer-range (greater than ~1 kb) interactions. These analyses identified two protein-protein crosslinkers that appeared to more efficiently crosslink distant nucleosomes within the *MDJ1* CID and to more efficiently capture CEN-CEN interactions – disuccinimidyl glutarate (DSG, a 7.7Å crosslinker), and ethylene glycol bis(succinimidyl succinate) (EGS, a 16.1Å crosslinker) (**Figure 3.1E**). The improvements in signal-to-noise afforded by DSG and EGS were not observed when DSG or EGS were added prior to cell permabilization (not shown), consistent with our expectation that these molecules

are too large to cross the yeast cell wall (SCHERRER *et al.* 1974). Interestingly, improved signal-to-noise was not observed following crosslinking with higher concentrations of formaldehyde, longer incubation times, or when a second round of formaldehyde crosslinking was carried out after cell wall digestion (**Figure 3.1B-D**), demonstrating that the improvements in the Micro-C protocol required some specific aspect of the DSG and EGS crosslinkers, rather than, say, an increase in the sheer density of crosslinks introduced into chromatin.

We incorporated each of these longer crosslinkers into an altered Micro-C protocol, which we dubbed Micro-C XL (MICROcoccal nuclease-based analysis of Chromosome folding using long X-Linkers), and then sought to identify those features of yeast chromosome folding uniquely revealed using these crosslinkers (**Figure 3.1F**). Briefly, actively growing budding yeast cultures are crosslinked with formaldehyde alone, formaldehyde + DSG, formaldehyde + EGS, or all three crosslinkers, and resulting chromatin is fragmented to mononucleosomes using MNase digestion. Crosslinked chromatin is then treated with T4 DNA polymerase in the absence of dNTPs to promote exonuclease activity. This leaves single stranded DNA ends, which are then repaired and biotinylated upon the addition of dNTPs, including biotin-dATP and biotin-dCTP, to the T4 polymerase reaction. Following DNA ligation, ligation products are purified away from unligated mononucleosomal DNA based 1) on ligation-dependent protection of biotinylated DNA from exonuclease attack, and 2) on size selection specifically of dinucleosome-sized ligation products. In practice, nucleosomal ligation products

are first treated with exonuclease III to remove biotinylated nucleotides from free DNA ends, leaving biotinylated nucleotides specifically in nucleosomal ends that had been ligated to one another and thereby protected from exonuclease attack. DNA is then purified from deproteinated chromatin, and dinucleosome-sized ligation products are gel-purified away from unligated mononucleosomal DNA. Recovered DNA is then further purified on streptavidin beads to isolate only DNA carrying biotinylated nucleotides at ligation junctions that had been protected from exonuclease digestion. Purified ligation products are then used to generate deep sequencing libraries, and subject to Illumina paired-end sequencing using standard methods.


A
S. cerevisiae
TEL
CEN
E
DSG (7.7 A)
EGS (16.1 A)
B
ChrVI: 100,252-106,236
MDJ1
LPD1
C
Viewpoint: Nuc#1
Viewpoint: Nuc#3
Viewpoint: Nuc#7
Ralative contact frequency (ratio of n/n+1)
FA1%
FA3%
DSG
EGS
NucA
NucB
NucC
NucD
D
Viewpoint: Nuc#1
Viewpoint: ChrIII:113,552
FA3%, 15min
FA3%, 45min
FA6%, 15min
FA6%, 45min
FA3%+FA0.5%
FA3%+FA1%
FA3%+DSG
CEN3-CEN7
CEN3-CEN16

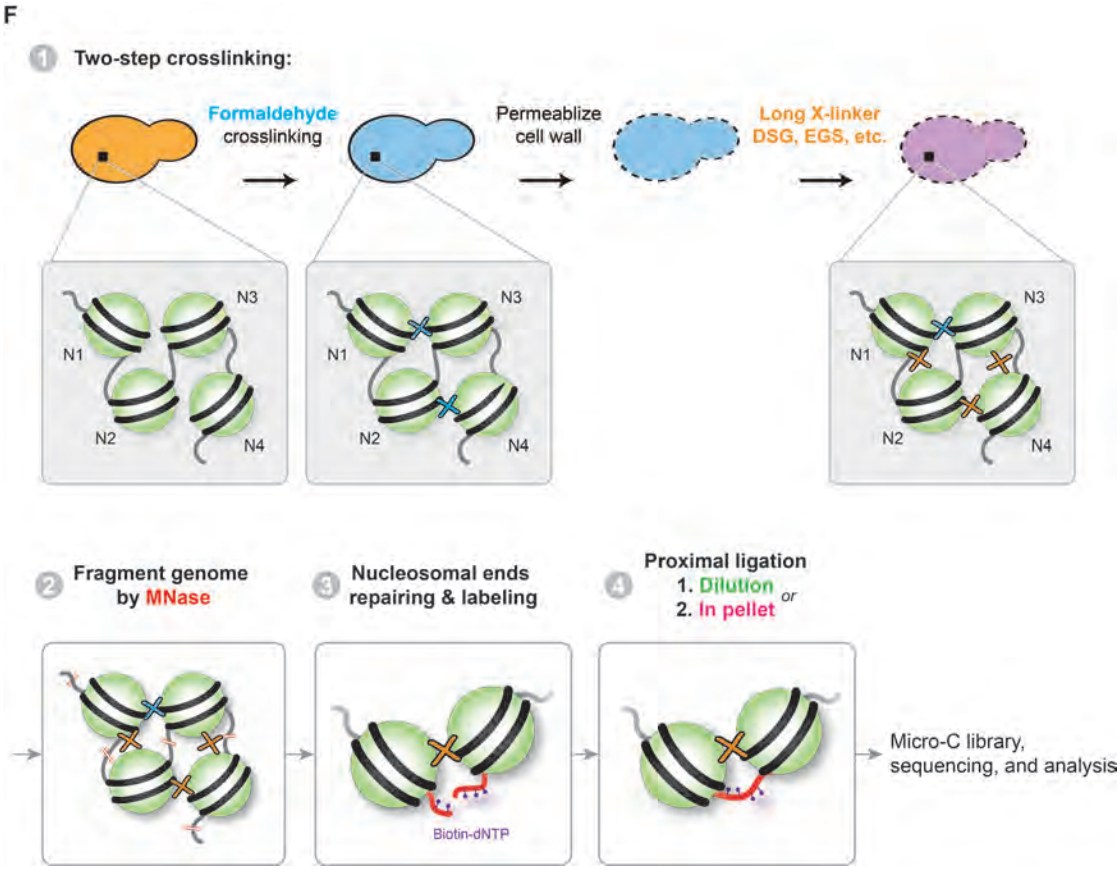


Figure 3.1. Overview of Micro-C XL.

(**A**) Schematic of global chromosomal organization in *S. cerevisiae*, highlighting the CEN-CEN cluster and the TEL-TEL clusters. (**B-D**) Q-PCR analysis of effects of long crosslinkers on Micro-C protocol. Budding yeast were crosslinked with formaldehyde, permeabilized, and then treated with one of several alternative crosslinkers. Mononucleosomal DNA was then processed using the Micro-C protocol, and ligated DNA was subject to q-PCR using primer pairs designed against a variety of nucleosomes surrounding the *MDJ1* contact domain or spanning specific centromere pairs. In (**B**), schematic of primer locations for q-PCR analyses of the ~6 kb region surrounding *MDJ1*. Forward primers located at nucleosomes 1, 3, and 7 are indicated in blue, while locations of reverse primers are shown in orange. In (**C**), q-PCR data for the indicated forward and reverse primer pairs, normalized to the q-PCR signal obtained for the abundance of ligation products between the upstream nucleosome in question and its immediate downstream neighbor (e.g., for the left panel, data are normalized to the

pairwise interaction between *MDJ1* nucleosomes number 3 and 4). Data are shown for 1% FA, 3% FA, 3% FA + DSG, and 3%FA + EGS, as indicated. Here (and throughout this study, with the exception of **Figure 3.1D**), FA crosslinking was carried out prior to cell permeabilization, and DSG or EGS crosslinking was introduced following cell permeabilization. In (**D**), (Left panel) As in (**C**), but showing data for additional crosslinking conditions including higher FA concentrations, longer FA crosslinking, and a two-step FA crosslinking protocol in which a second FA incubation is carried out after spheroplasting. This last protocol mimics the use of DSG or EGS after spheroplasting in the Micro-C XL protocol. (Right panel) q-PCR data for interactions between *CEN3* and the indicated centromeres, showing data for the same protocols detailed in the left panel. (**E**) Structures of the two protein-protein crosslinkers used in Micro-C XL. (**F**) Outline of changes to the Micro-C protocol. After budding yeast are fixed with formaldehyde, cells are permeabilized, then treated with one of several additional protein-protein crosslinkers. Crosslinked chromatin is then digested to mononucleosomes using micrococcal nuclease. End digestion and repair is used to introduce biotinylated nucleotides into mononucleosomal ends, and nucleosomes crosslinked to one another are ligated together at high dilution or “in pellet”. Ligation products are then purified via streptavidin capture and size selection of dinucleosome-sized DNA, and paired-end deep sequencing is used to characterize internucleosomal interactions genome-wide.

Genome-wide analysis of chromosome folding by Micro-C XL

To investigate whether adding DSG or EGS to the Micro-C protocol provided additional insights into chromosome folding, we generated genome-wide Micro-C maps for budding yeast subject to a variety of crosslinking conditions. These conditions include 1% or 3% formaldehyde (FA) alone, or FA (1% or 3%) plus DSG, EGS, or both crosslinkers. Also, as described above, we generated similar datasets in which DSG or EGS were added prior to cell wall digestion as a negative control, as these molecules are not expected to cross the cell wall in

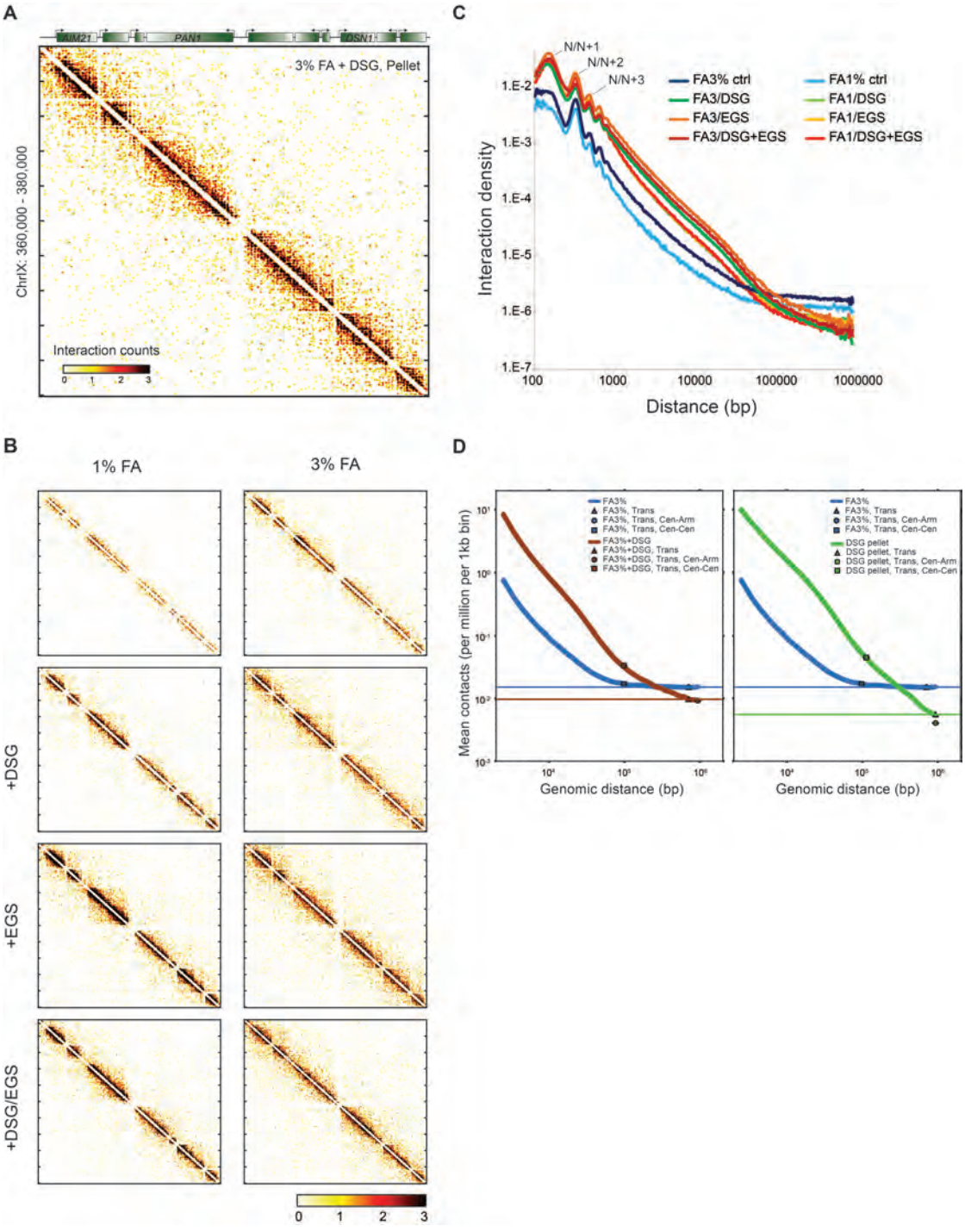
budding yeast. Below, we primarily focus on the results using 3% FA with or without DSG and EGS, but results of other crosslinking conditions are noted when relevant.

In general, all four conditions (FA, FA/DSG, FS/EGS, and FA/DSG/EGS) yielded qualitatively similar results at the scale of individual genes, with chromosomal interaction domains of varying strength covering ~1-5 genes (**Figure 3.2A**). Although CIDs were clearly observed in all four conditions, the addition of longer crosslinkers to the Micro-C protocol resulted in improved ability to visualize these structures (**Figure 3.2B** and see below). Importantly, as previously observed with Micro-C, we again found no evidence for a regular organization of the chromatin fiber above the nucleosomal scale, which would have manifested as a peak in interaction density at a genomic distance corresponding to the fiber size (**Figure 3.2C**). Moreover, compared to standard Micro-C we found that Micro-C XL exhibited substantially higher signal-to-noise (**Figure 3.2D**), consistent with the q-PCR results in **Figure 3.1B-D**.

Beyond recapitulating the key aspects of chromosome folding previously revealed by Micro-C, Micro-C XL resolved additional details that were not apparent in prior Micro-C maps. Most interestingly, in contrast to standard Micro-C crosslinking conditions, all three long crosslinking conditions captured very robust CEN-CEN and TEL-TEL interactions characteristic of the Rabl configuration for interphase chromosomes (DUAN *et al.* 2010; MARBOUTY *et al.* 2014; MIZUGUCHI *et al.* 2014) (**Figures 3.2E-F**). This finding thus resolves the

primary qualitative discrepancy between prior Micro-C data and known features of genomic folding in yeast, while preserving the ability of Micro-C to interrogate chromatin interactions at the 2-10 nucleosome scale.

We conclude that both DSG and EGS dramatically extend the length scale at which chromosome folding can be assayed by Micro-C, enabling analysis at scales from the local chromatin fiber to the full genome. Interestingly, this improvement can largely be ascribed to a decrease in the background levels of ligation between distant genomic regions (**Figure 3.2D**, left panel) – the decrease in this “noise floor” seen using the Micro-C XL protocol is likely to account for the improved ability to measure relatively low-abundance signals such as CEN-CEN interactions.


A
VIM21
PAN1
DSN1
3% FA + DSG, Pellet
ChrIX: 360,000 - 380,000
Interaction counts
0 1 2 3
C
N/N+1
N/N+2
N/N+3
FA3% ctrl
FA3/DSG
FA3/EGS
FA3/DSG+EGS
FA1% ctrl
FA1/DSG
FA1/EGS
FA1/DSG+EGS
Interaction density
1.E-2
1.E-3
1.E-4
1.E-5
1.E-6
1.E-7
100
1000
10000
100000
1000000
Distance (bp)
B
1% FA
3% FA
+DSG
+EGS
+DSG/EGS
0 1 2 3
D
Mean contacts (per million per 1kb bin)
FA3%
FA3%, Trans
FA3%, Trans, Cen-Arm
FA3%, Trans, Cen-Cen
FA3%+DSG
FA3%+DSG, Trans
FA3%+DSG, Trans, Cen-Arm
FA3%+DSG, Trans, Cen-Cen
DSG pellet
DSG pellet, Trans
DSG pellet, Trans, Cen-Arm
DSG pellet, Trans, Cen-Cen
Genomic distance (bp)
Genomic distance (bp)

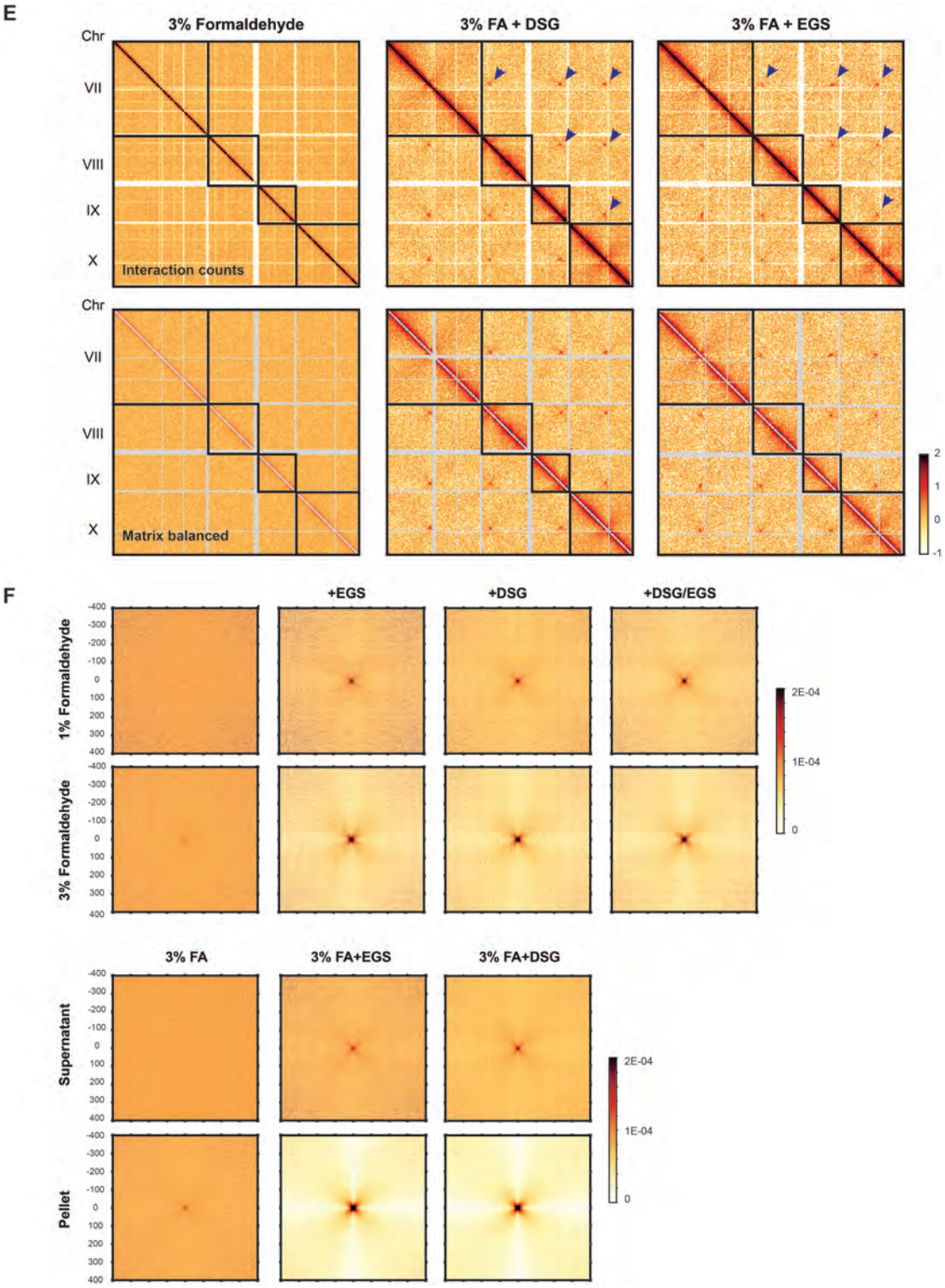

E
3% Formaldehyde
3% FA + DSG
3% FA + EGS
Chr
VII
VIII
IX
X
Interaction counts
Matrix balanced
2
1
0
-1
F
+EGS
+DSG
+DSG/EGS
1% Formaldehyde
3% Formaldehyde
2E-04
1E-04
0
3% FA
3% FA+EGS
3% FA+DSG
Supernatant
Pellet

Figure 3.2. Micro-C XL robustly captures known interchromosomal interactions while retaining single-nucleosome resolution.

(**A**) Example of Micro-C XL contact map for a 20 kb genomic stretch at single-nucleosome resolution. The map is shown for ChrIX: 360,000-380,000 for crosslinking condition in 3% formaldehyde + DSG and ligation in pellet. The raw matrix was only normalized to sequencing depth as described in (HSIEH *et al.* 2015), and interactions were counted in single bp resolution without binning. The Micro-C XL protocol effectively recovers the chromosomally-interacting domains previously observed in Micro-C data. (**B**) Comparison of crosslinking protocols for a typical 20 kb region. Micro-C data are shown for ChrIX: 360,000-380,000 for eight different crosslinking conditions, as indicated. The raw matrix was only normalized to sequencing depth as described in (**A**) and (HSIEH *et al.* 2015). Improved capture of contact domains associated with individual genes is readily apparent here for protocols incorporating DSG or EGS. (**C**) Plot of interaction density for all unidirectional ("IN-OUT") read pairs, expressed as a fraction of potential pairwise interactions (per bp^2) (y axis, $\log_{10}$), vs. genomic distance (x axis, $\log_{10}$) for various Micro-C protocols scaled to 10^9 interactions. (**D**) Addition of long crosslinkers reduces the "noise floor" relative to FA-only Micro-C maps. Intra-arm contact probability, *P(s)*, as a function of genomic distance, s, was calculated from 1 kb corrected contact maps as in (IMAKAEV *et al.* 2015), using 50 logarithmically spaced bins from 1kb to 1Mb. Horizontal line marks average trans (between-chromosome) contact frequency. Markers respectively indicate average trans, trans Cen-vs-Cen and trans Cen-vs-Arm contact frequencies, defining each centromere with a +/-20kb genomic window, and are placed at the genomic distance with an equivalent intra-arm contact probability for ease of comparison. Note that *P(s)* flattens out at the average trans contact frequency in the FA3-only dataset, as would be result from an adding a constant frequency of interaction between any two loci. Additionally, while trans-cen-arm and trans-average are similarly strong in the FA3-only and FA3-DSG datasets, the avoidance of the centromere from arm regions is clearly seen in the DSG-pellet dataset. This decrease in the background levels of ligation between distant genomic regions – the "noise floor" – seen using the Micro-C XL protocol is likely to account for the improved ability to measure relatively low-abundance signals such as CEN-CEN interactions. (**E**) Interaction maps for Micro-C data generated using 3% formaldehyde, 3% formaldehyde

+ DSG, or 3% formaldehyde + EGS are shown as indicated for budding yeast chromosomes VII through X. Here, data on the top panel are normalized by read depth only and displayed as $\log_{10}$ (counts per million) and data on the bottom panel are further corrected by matrix balance. Note that since genomes were fragmented more homogeneously (mononucleosome) in Micro-C protocol than in Hi-C, matrix correction is not always necessary for data visualization in a large size binning (> 1kb), as shown in here. For 3% FA + DSG and 3% FA + EGS, CEN-CEN interactions are shown with blue arrows. As previously observed, telomere-telomere interactions are only observed between a subset of chromosome arms (which do not include interactions between chromosomes VII to X) in budding yeast. (**F**) Centromere clustering revealed by alternative crosslinkers. Average interaction map for all possible pairs of CEN-CEN interactions for the indicated Micro-C protocols. Top eight panels show variants of the standard Micro-C protocol performed using the indicated crosslinking conditions, while bottom six panels show data for Micro-C performed following separation of relatively soluble and insoluble MNase-digested chromatin by centrifugation prior to ligation. 1 kb binned contact maps were corrected for genomic coverage and normalized such that the total coverage of each 1 kb region summed to 1.

Bona-fide Micro-C contacts are primarily found in relatively insoluble chromatin

We next sought to uncover whether Micro-C data are affected by fractionation of crosslinked chromatin prior to proximity ligation. This was motivated by the absence of “gene loops” (O'SULLIVAN *et al.* 2004; ANSARI AND HAMPSEY 2005) in our previous Micro-C analysis – the 3C method used in several studies of gene loops includes a step in which insoluble chromatin is pelleted and isolated prior to ligation, and it is known that different chromatin structures are likely to be present in soluble vs. insoluble crosslinked chromatin (HENIKOFF *et al.* 2009). We

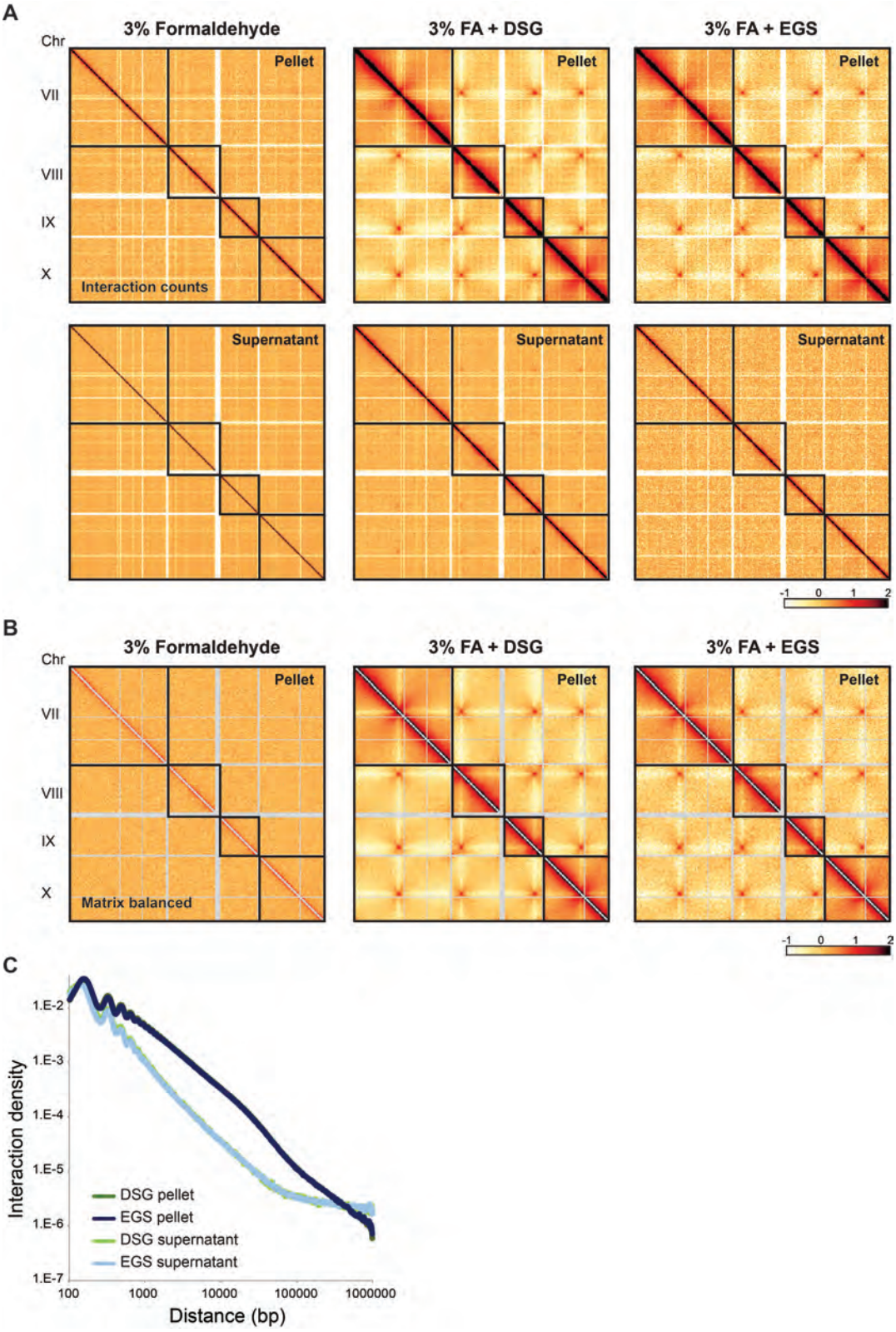
therefore carried out Micro-C XL in which fragmented chromatin was centrifuged after the completion of MNase digestion to separate soluble from insoluble chromatin (**Figure 3.1F** and **Methods**), and proximity ligation was carried out separately on pellet and supernatant material (**Figure 3.3**).

Micro-C XL maps from supernatant material were extremely noisy at longer distances, and did not identify known aspects of higher-order organization. In contrast, contact maps generated from relatively insoluble chromatin had excellent signal to noise, and robustly captured CEN-CEN interactions. We conclude that this reflects either preferential precipitation of crosslinked fragments, higher ligation efficiency in the pellet, or higher specificity of ligation in the pellet; i.e. noise in the supernatant dataset is elevated due to ligations in solution between freely-moving, likely uncrosslinked, nucleosomes causing artefactual contacts in trans. These related hypotheses are of course not mutually exclusive, and may all contribute to the improved signal to noise seen in Micro-C XL maps from pellet material (**Figure 3.2D**, right panel).

We next searched for evidence of gene loops (O'SULLIVAN *et al.* 2004; ANSARI AND HAMPSEY 2005) in the dataset generated from relatively insoluble chromatin. Here, we consider a gene “loop” to be characterized by an increased contact frequency between the gene start and stop relative to other locus pairs in their vicinity (similar to previous definitions of peaks in Hi-C contact maps (RAO *et al.* 2014)), rather than a gene-wide increase in relative contact frequency. In general, it is clear that, for any given nucleosome, raw interaction counts (either

normalized only for library depth, or normalized additionally for nucleosome occupancy) decay steadily with increasing distance and do not exhibit an uptick at gene ends, which is the signature of a looping interaction (BENEDETTI *et al.* 2014; DOYLE *et al.* 2014) – see example in **Figure 3.2A**, or averaged "metagenes" in **Figures 3.3D-F**. Nevertheless, with our population-average contact maps we cannot rule out a scenario where populations of various length gene loops are formed dynamically over gene bodies (as proposed for enhancer-promoter interactions (LEE *et al.* 2015)).

Although our data thus do not support the concept of widespread end-to-end gene loops at transcribed genes, visual inspection of Micro-C XL data did reveal a small number of possible looping interactions at multi-gene scale that were apparent even in interaction counts not normalized for distance (not shown). Validation and functional analysis of these apparent loops will be the subject of future studies.


A
3% Formaldehyde
3% FA + DSG
3% FA + EGS
Chr
VII
VIII
IX
X
Pellet
Supernatant
Interaction counts
-1 0 1 2
B
Matrix balanced
C
Interaction density
1.E-2
1.E-3
1.E-4
1.E-5
1.E-6
1.E-7
DSG pellet
EGS pellet
DSG supernatant
EGS supernatant
100
1000
10000
100000
1000000
Distance (bp)

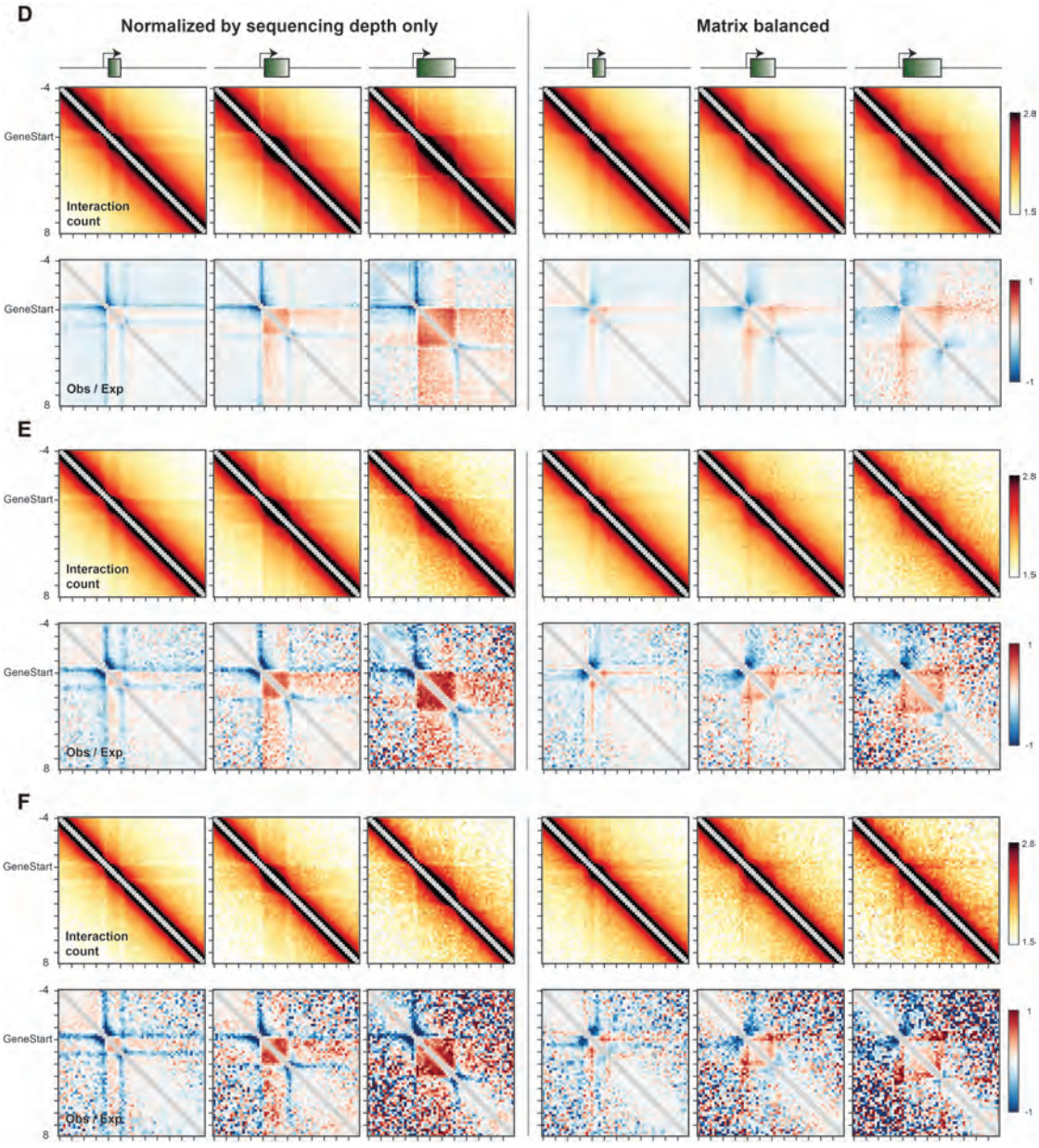


Figure 3.3. Micro-C XL interactions are enriched in insoluble chromatin.

(**A**) Micro-C interaction maps for chromosomes VII to X are shown as in **Figure 3.2**, for pellet (top panels) and supernatant (bottom panels) libraries as indicated. Note that the strong enrichment of CEN-CEN interactions in the pellet fraction requires long-distance crosslinkers, as it is not clearly observed for 3% FA chromatin pellets (also see **Figure 3.2F**, bottom panels). (**B**) Here are shown the same data as in (**A**, top panels) for further correction by matrix balancing. (**C**) Plot of interaction density (y axis, $\log_{10}$) vs. genomic

distance (x axis, $\log_{10}$) for four Micro-C XL libraries, normalized as in **Figure 3.2C**. One pair of libraries were crosslinked with 3% FA + DSG, then MNase-digested chromatin was separated into soluble and insoluble fractions by centrifugation; the same procedure was also repeated for yeast crosslinked with 3% FA + EGS. In both cases, the relatively soluble Micro-C library exhibited far lower signal to noise, with relatively rapid decay of interactions with increasing distance, compared to libraries constructed from pellet material (also see **Figure 3.2D**, right panel). (**D-F**) Micro-C XL pellet metagenes display no evidence for widespread gene loops. As prior reports of gene loops in budding yeast were based on 3C data generated from insoluble chromatin, we sought evidence of such loops in the Micro-C XL dataset generated from relatively insoluble chromatin. Similar to previous definitions of a loop, we consider a "gene loop" to be characterized by an increased contact frequency between the gene start and stop relative to other locus pairs in their vicinity, rather than a gene-wide increase in relative contact frequency. In general, raw interaction counts for any given nucleosome (either normalized only for library depth, or normalized additionally for nucleosome occupancy) decay steadily with increasing distance and do not exhibit an uptick at gene ends, which is the signature of a looping interaction. Similarly, visual inspection of Micro-C XL maps does not reveal any evident peaks of contact probability between the 5' and 3' ends of genes (e.g., see **Figures 3.2A-B**). In (**D**, left panels), metagene maps for *S. cerevisiae* DSG pellet dataset, binned to 200 bp resolution and normalized only for sequencing depth. All genes of length 1-1.2 kb, 2-2.2 kb, and 3-3.2 kb, as indicated, were identified and aligned by their 5' ends. The narrow range of gene lengths was chosen to assist in visualization of a discrete 3' gene end in these plots. Top panels show $\log_{10}$ averaged interaction counts, normalized only for library read depth. CID structure is evident in these panels as a region of increased contacts bounded at both the 5' and the 3' ends of genes. Note that interactions within each box decay smoothly with increasing distance from the diagonal, indicating that interactions between gene ends are at most a minority subpopulation of gene folding conformations. Bottom panels show the same data, after additionally controlling for the global decay in interaction frequency with increasing genomic distance. Data are shown as $\log_2$ of the observed interactions divided by the interaction count expected based on genomic distance. This correction reveals a far clearer view of CID structure, with clear blue boundaries delimiting the red contact

domain associated with the gene. In (**D**, right panels), as in left panels, but data are additionally normalized by matrix balancing, which corrects for possible experimental biases as well as nucleosome occupancy (observed as variation in the total coverage per bin in Micro-C maps). Visually this removes the faint “stripes” in the raw data (top panels) associated with nucleosome-depleted promoter regions, but reveals red stripes of excess observed/expected interactions (bottom panels) for +1 and +N nucleosomes. Following this normalization a subtle enrichment of interactions can be observed for the +1/+N nucleosome interaction in the observed/expected visualization (bottom row). Regardless of row normalization scheme, we do not observe the signature of a gene loop, an uptick in contact probability between the +1/+N nucleosomes, in observed maps for any of the assayed gene lengths. With regards to observed/expected normalization, we note that even uniform squares on an observed map can have apparent corner peaks after dividing by an expected map where contact frequency decreases with genomic distance. For these reasons, Micro-C data argues for the prevalence of gene- and multi-gene-wide crumpling, rather than specific +1/+N gene loops. Although our data thus do not support the concept of widespread end-to-end gene loops at transcribed genes, visual inspection of Micro-C XL data did reveal a small number of possible looping interactions at multi-gene scale that were apparent even in interaction counts not normalized for distance (now shown). Validation and functional analysis of these apparent loops will be the subject of future studies. Metagene visualization as in (**D**) for (**E**) 3% FA + DSG without spin down, and (**F**) 3% FA, Micro-C libraries.

Comparison of chromosome folding in *S. cerevisiae* and *S. pombe*

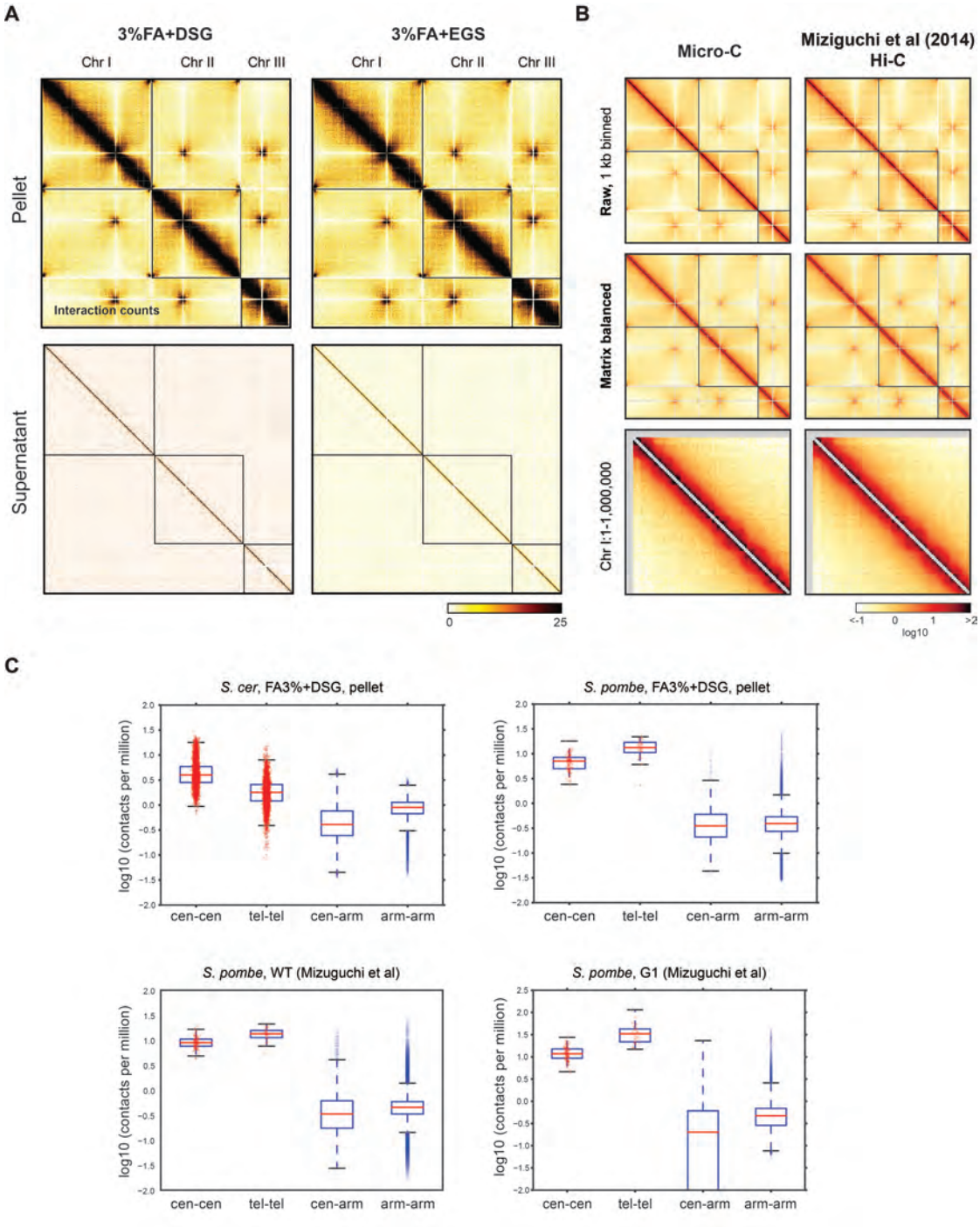
Although many aspects of chromosome folding are conserved between *S. cerevisiae* and other eukaryotes, *S. cerevisiae* lacks several evolutionarily widespread chromatin regulatory systems, such as the H3K9me3/HP1 and H3K27me3/Polycomb systems for gene repression found in many eukaryotes. We therefore carried out Micro-C XL in the fission yeast *S. pombe* to ascertain the similarities and differences in chromosome folding between these distantly-

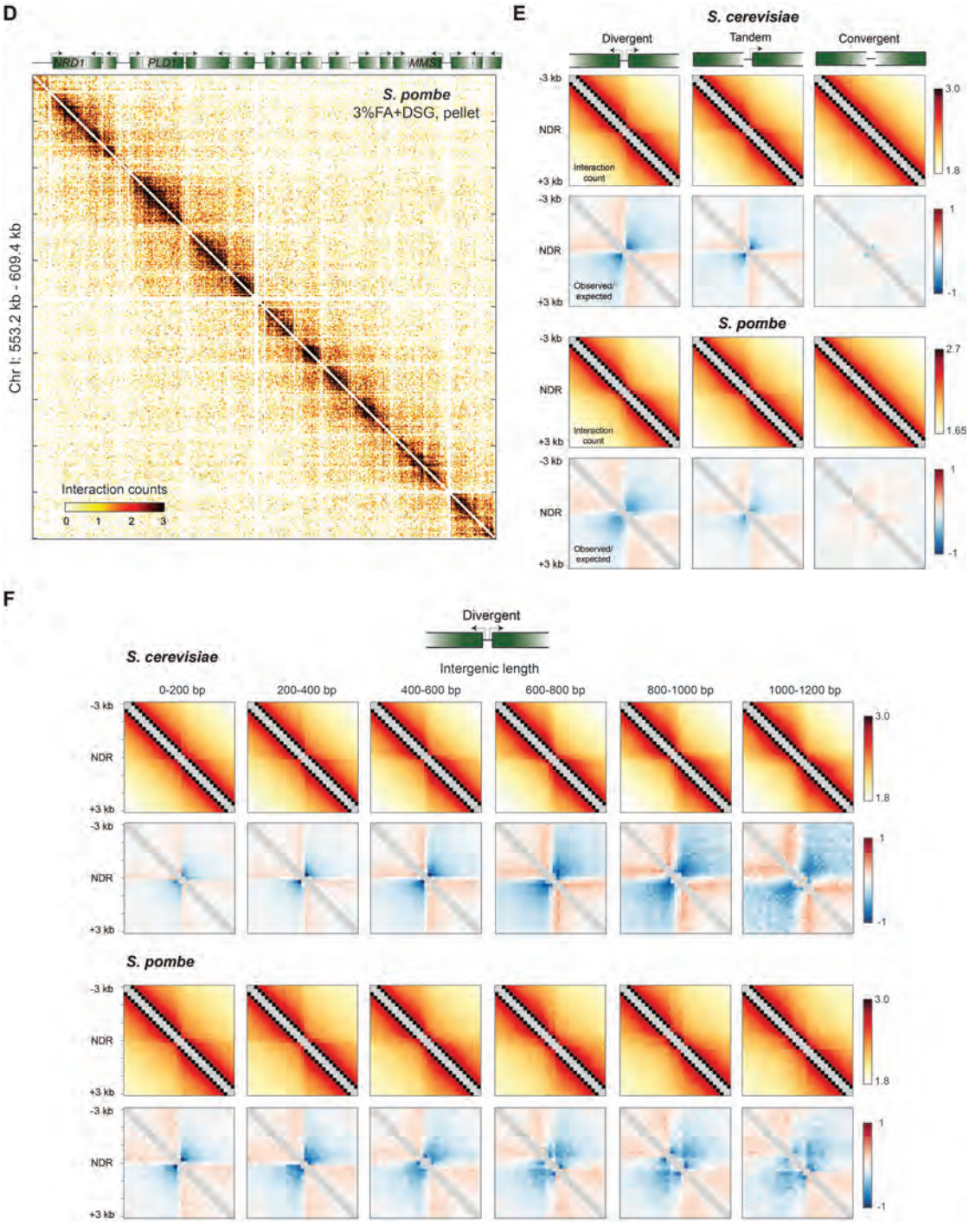
related microbes, and to demonstrate the broad applicability of our methods. Key aspects of the Micro-C XL protocol proved equally important in fission yeast, as for example maps generated from relatively insoluble chromatin exhibited far less noise compared to maps based on soluble chromatin (**Figure 3.4A**). Overall, our data were well-correlated (Spearman's r = 0.77 using 10 kb bins) with a prior Hi-C analysis of *S. pombe* chromatin by (MIZUGUCHI *et al.* 2014) (**Figure 3.4B**). As in budding yeast, Micro-C XL maps in *S. pombe* revealed frequent interactions along the diagonal, robust CEN-CEN and TEL-TEL interactions, and a depletion of interactions between centromeres and chromosome arms (**Figures 3.4A-B**). We did find quantitative differences in such large-scale aspects of chromosome folding, as *S. pombe* chromosomes exhibited slightly stronger centromere clustering, and substantially stronger telomere clustering (**Figure 3.4C**). These differences do not appear to be a consequence of the profound cell cycle differences between budding and fission yeast (**Figure 3.4C**), but could potentially be explained by any number of other features ranging from the smaller number of longer chromosomes in *S. pombe*, to the molecular details of interactions between pairs of H3K9-methylated nucleosomes present in this species but not in *S. cerevisiae*.

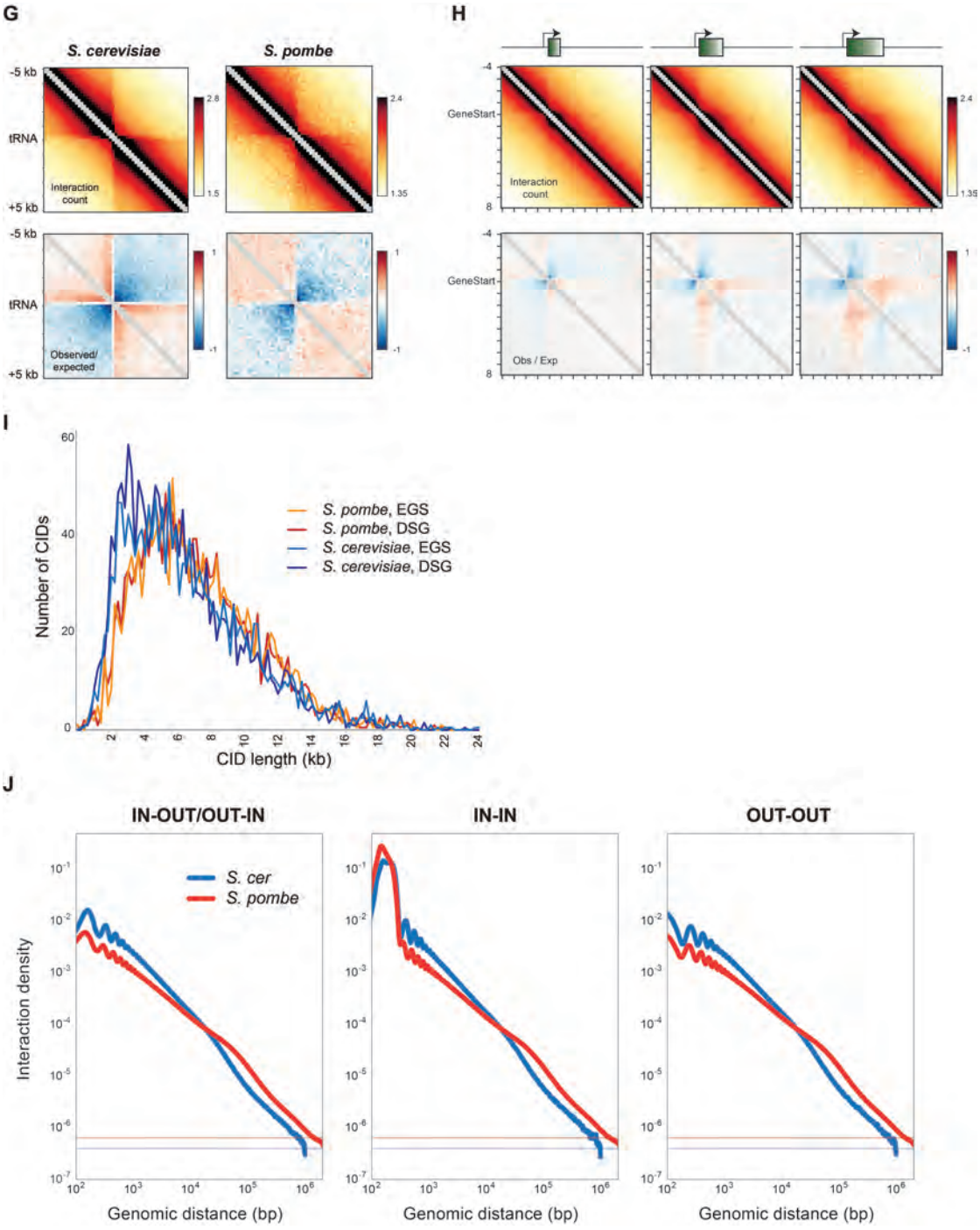
As prior studies of chromosome folding in *S. pombe* were performed with ~10 kb kb resolution, we next turned to those aspects of chromatin structure uniquely interrogated using the enhanced resolution of Micro-C. Visual inspection of chromosome folding revealed abundant contact domains associated with ~1-5

genes and separated by promoter regions (**Figure 3.4D**), analogous to the CID structures in budding yeast. As in budding yeast, promoters in fission yeast acted as efficient boundaries between CIDs, and metagene analysis revealed remarkably similar behavior in both yeast species at the length scale of individual genes or promoters (**Figures 3.4E-H**). In addition, plots of interaction frequency vs. distance, and distributions of contact domain length, were qualitatively similar, yet quantitatively different, in these two species; in particular, differences at short distances in the positions of interaction maxima correspond to known differences in nucleosome repeat length in these two species (**Figure 3.4I**).

We conclude that broadly similar principles underlie chromosome folding behavior in these distantly-related fungi, with modest quantitative differences in chromosome structure that could potentially result from the interspecies differences in aspects of genomic structure including chromosome length, gene length, intron abundance, location of rDNA clusters, and nucleosome repeat length.


A
3%FA+DSG
3%FA+EGS
Chr I
Chr II
Chr III
Pellet
Supernatant
Interaction counts
0
25
B
Micro-C
Miziguchi et al (2014)
Hi-C
Raw, 1 kb binned
Matrix balanced
Chr I:1-1,000,000
<-1
0
1
>2
log10
C
S. cer, FA3%+DSG, pellet
S. pombe, FA3%+DSG, pellet
S. pombe, WT (Mizuguchi et al)
S. pombe, G1 (Mizuguchi et al)
log10 (contacts per million)
cen-cen
tel-tel
cen-arm
arm-arm


D
NRD1
PLD1
MMS1
S. pombe
3%FA+DSG, pellet
Chr I: 553.2 kb - 609.4 kb
Interaction counts
0 1 2 3
E
S. cerevisiae
Divergent
Tandem
Convergent
-3 kb
NDR
+3 kb
Interaction count
Observed/ expected
3.0
1.8
1
-1
S. pombe
2.7
1.65
F
Divergent
S. cerevisiae
Intergenic length
0-200 bp
200-400 bp
400-600 bp
600-800 bp
800-1000 bp
1000-1200 bp
S. pombe


G
S. cerevisiae
S. pombe
-5 kb
tRNA
+5 kb
Interaction count
Observed/ expected
H
GeneStart
Interaction count
Obs / Exp
I
Number of CIDs
CID length (kb)
S. pombe, EGS
S. pombe, DSG
S. cerevisiae, EGS
S. cerevisiae, DSG
J
IN-OUT/OUT-IN
IN-IN
OUT-OUT
S. cer
S. pombe
Interaction density
Genomic distance (bp)

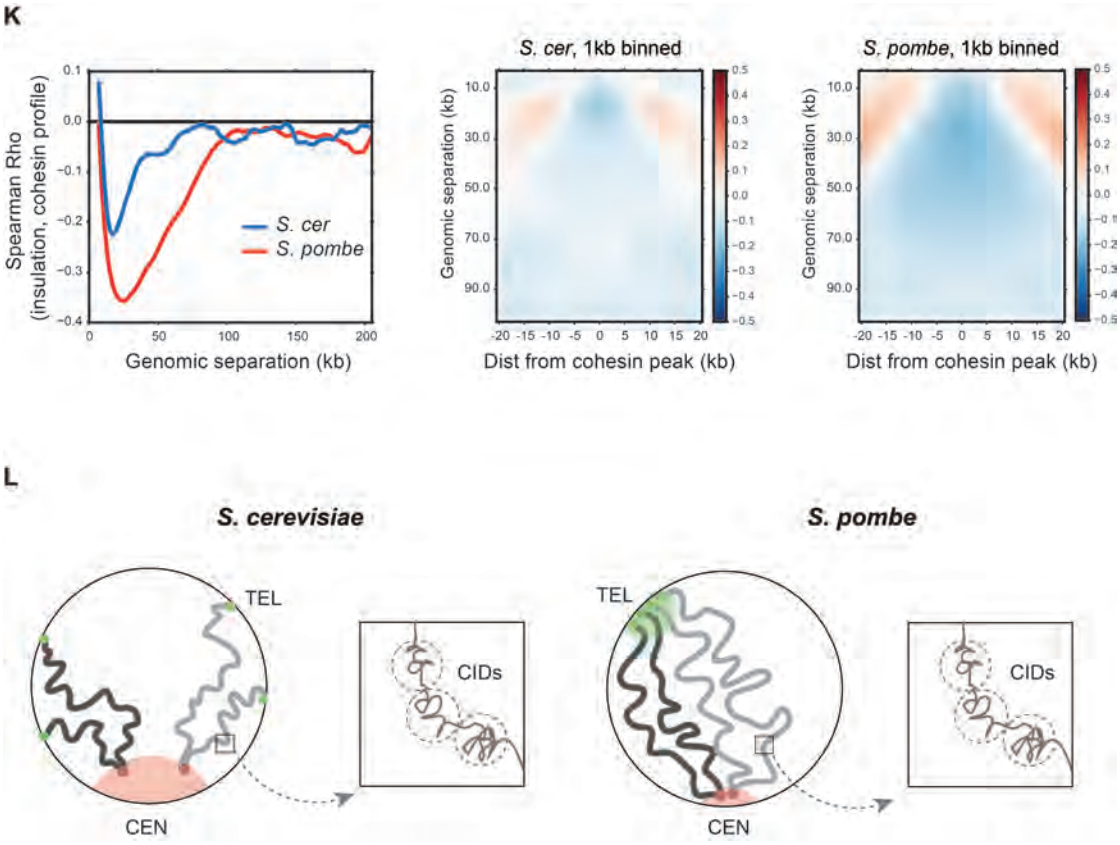


Figure 3.4. Comparative analysis of chromosome folding in *S. cerevisiae* and *S. pombe*.

(**A**) Whole genome Micro-C XL interaction map for *S. pombe*. Here, data are shown for yeast crosslinked with 3% FA + DSG (left panels) and 3% FA + EGS (right panels), and pelleted prior to ligation (insoluble and soluble materials shown as indicated). Raw matrixes were normalized to sequencing depth only and scaled to 10^9 reads, and interactions were counted with 5 kb binning. Key features of this map include robust clustering of centromeres and telomeres (with the exception of the rDNA-carrying chromosome III telomeres, which were excluded from analysis based on their repetitive nature), and strong depletion of interactions between centromeres and chromosome arms. Those features are not observed in the supernatant data. (**B**) Comparison of Micro-C and published Hi-C maps for *S. pombe* at 10 kb resolution. In each row, left panels show data this study, while right panels are from (MIZUGUCHI *et al.* 2014), binned at 10 kb resolution in $\log_{10}$ count. Top two rows show data for the entire genome, while bottom row shows a 1 MB zoom-in. Overall Micro-C maps are highly-correlated with

published results for this species (Spearman's r = 0.77 for corrected 10 kb resolution maps, comparable with the r = 0.77 correlation between DSG and EGS pellet maps for *S. cerevisiae*), with both maps showing ~100 kb chromatin contact domains previously referred to as cohesin-mediated "globules". (**C**) Comparison of global folding behavior between two yeast species. Boxplots showing fraction of contacts (normalized as parts per million read pairs) for interactions between centromeres (cen-cen), between telomeres (tel-tel), between centromeres and chromosome arms (cen-arm), and between distal chromosome arms (arm-arm), calculated from coverage-corrected 10 kb binned contact maps. Chromosome arms here are defined as sequences more than 20 kb away from either a centromere or telomere. Boxplots here show the median (red line), the first and third quartiles (box), and 1.5* the inner quartile range (whiskers). Red points overlay values for bin-pairs for (cen-cen) and (tel-tel) regions. As in (MIZUGUCHI *et al.* 2014), the 10 most telomere-proximal bin-pairs for non-filtered regions of the heatmap are chosen, or 40 most centromere-proximal bin pairs (as there are 4 arm pairs at each centromere). Data for upper panels are taken from this study, while lower panels show data from (MIZUGUCHI *et al.* 2014) for unsynchronized or G1-arrested *S. pombe* (lower left and right, respectively). Note that the enhanced telomere clustering seen in *S. pombe* relative to *S. cerevisiae* is observed both using Micro-C (upper right panel) and Hi-C (lower left panel) in unsychronized *S. pombe*, and in G1-arrested (lower right panel) *S. pombe*, indicating that this difference between budding and fission yeast does not result from the profound cell cycle differences between budding and fission yeast. Instead, these quantitative differences in chromosome packing might potentially be explained by any number of other features ranging from the smaller number of longer chromosomes in *S. pombe*, to the molecular details of interactions between pairs of H3K9-methylated nucleosomes that are present in this species but not in *S. cerevisiae*. (**D**) Zoom-in on *S. pombe* Chr I: 553,200-609,400, showing widespread contact domains typically associated with individual genes, but occasionally associated with blocks of ~2-5 genes. Matrix was counted by pairs per million in single bp-resolution and normalized to sequencing depth only. (**E**) Comparison of chromosome folding in S. cerevisiae and S. pombe. For each species, intergenic regions were separated into those falling between pairs of genes oriented divergently, in tandem, or convergently, as indicated, and were aligned according to the midpoint of the nearest respective intergenic region.

Data from DSG pellet libraries from *S. cerevisiae* (upper panels) or *S. pombe* (lower panels) are averaged for all genes in each category. For both species, two rows of panels are shown as in **Figures 3.3D-F**, with interaction counts in top panels and distance-corrected interaction levels in bottom panels. Interaction counts were normalized to library read depth for coverage-corrected 200 bp-binned maps for the three classes of intergenic region. Distance-corrected maps are the same data, additionally corrected for the decay in interaction frequency with increasing distance, and expressed as the $\log_2$ ratio of observed interactions divided by expected interactions for a given genomic distance. (**F**) Comparison of boundary activity of promoters in budding and fission yeast. For both species, two rows of panels are shown as in (**E**), with interaction counts in top panels and distance-corrected interaction levels in bottom panels. In both cases data are from DSG pellet maps. Data here are shown for divergently-oriented genes, separated into groups based on the intergenic distance. In both species, divergent promoters act as boundaries, with longer intergenic regions more effectively separating chromatin domains from one another. Budding and fission yeast also exhibit similar behavior at intergenic regions separating tandemly-oriented genes, and separating convergently-transcribed genes (not shown). (**G**) tRNA metagene analysis. Data for *S. cerevisiae* and *S. pombe* DSG pellets are shown here aligned for all tRNA genes in the respective yeast genome. For both species, data shown here were normalized with matrix balancing. In each case, top panel shows interaction counts, while bottom panel shows observed/expected relative to interaction distance. (**H**) Metagene analysis of S. pombe genes. As in **Figures 3.3D-F**, for *S. pombe* DSG pellet data. As observed for *S. cerevisiae*, gene loops are not observed in coverage-corrected interaction data, but distance-corrected interactions reveal compacted domains at the gene level. (**I**) Distribution of contact domain lengths. Boundaries between contact domains were called as described in (HSIEH *et al.* 2015) for *S. cerevisiae* and *S. pombe* Micro-C XL datasets. Plots show the distribution of lengths for boundary-delimited contact domains, which are extremely similar for these two species. (**J**) Decay of Micro-C XL interactions with increasing genomic distance. Interactions vs. distance are shown for the indicated read pair orientations for the two species. Subtle differences at short distances are primarily attributable to different nucleosome repeat lengths in these species, while at longer distances we find *S. pombe* interactions decay slightly more

slowly with increasing distance. Interestingly, we also note an inflection point at ~80 kb at which interactions in *S. pombe* decay more rapidly – this may reflect the more robust organization of the fission yeast genome into cohesin-delimited "globules" (also see **Figure 3.4B** and below). (**K**) Cohesin insulates chromatin domains from one another. (Left panel) Correlation between cohesin localization and local chromatin insulation (y axis), at varying offsets (x axis). Here, insulation was calculated as using sliding diamond window, as in Sofueva *et al*. Insulation profiles were calculated from 1 kb binned and corrected DSG pellet contact maps, using a 10 kb sliding window at each indicated offset. Note that cohesin localization is correlated with the local insulation score in both budding and fission yeast, but that cohesin-associated insulation in fission yeast is far stronger (deeper peak), and extends over greater genomic distances (peak width). Cohesin localization for *S. cerevisiae* was obtained from GEO GSE42655 (Scc1) (ENERVALD *et al.* 2013), divided by input, $\log_2$-transformed, and binned to the same 1 kb resolution as contact maps. Cohesin for *S. pombe* was obtained from GEO GSE56848 (Psc3 WT) (MIZUGUCHI *et al.* 2014), $\log_2$-normalized by input, and binned to 1 kb. (Right panels) Average insulation profiles for Micro-C contacts surrounding cohesin binding sites in the indicated species. Cohesin peaks were called as local maxima on the binned 1 kb $\log_2$ profile, and were additionally required to have a minimum spacing of 10 kb and be in the top 75th percentile overall. Although both species exhibit local insulation, seen here as a blue depletion of contacts centered on cohesin binding sites, the inhibition of crossing interactions occurs at far greater distances (up to ~75 kb) in *S. pombe* than in *S. cerevisiae* (~25 kb). While displaying qualitative similarities, the quantitative differences captured here – insulation by cohesin-associated loci is stronger and persists over greater genomic distances in *S. pombe* relative to *S. cerevisiae* – may point to important differences in the underlying biology of cohesin in these two highly diverged yeast species. In particular, cohesin has been reported not to display peaks along the chromosomal arms in *S. cerevisiae* G1 (HU *et al.* 2015), whereas peaks of cohesin binding in *S. pombe* G1 have been reported to coincide with regions of local insulation in *S. pombe* G1 Hi-C maps (MIZUGUCHI *et al.* 2014). Together, these observations point towards a more important role for cohesin in organizing the arms of *S. pombe* chromosomes in interphase, relative to *S. cerevisiae*. (**L**) Cartoon models of budding and fission yeast chromosome folding. Fission yeast exhibit a subtle enhancement in

centromere clustering, as well as much more substantially-enhanced telomere clustering, but at the level of individual genes fission yeast chromosomes exhibit similar folding properties to budding yeast chromosomes.

DISCUSSION

Here, we present an improved protocol for nucleosome-resolution mapping of chromosome folding, termed Micro-C XL. The primary technical improvements detailed here are 1) the use of additional “long-range” crosslinkers to supplement formaldehyde crosslinking, and 2) fractionation of relatively insoluble chromatin prior to nucleosome ligation and subsequent library construction. Contrary to our initial expectations, the dramatic improvement seen in apparent capture of long-range interactions using these protocols likely results not from the ability of long-range crosslinkers to bridge interacting genomic loci associated with proteins that are more than 3 Å away from one another, but rather from a decrease in the noise caused by soluble nucleosomes encountering one another in solution during the ligation reaction and causing artefactual “interactions” between unlinked nucleosomes (**Figure 3.2D**). This hypothesis is based on the fact that DSG- and EGS-based Micro-C maps are extremely similar despite their substantial difference in crosslinking distance, as well as the finding that isolation of soluble chromatin results in greatly increased noise in Micro-C maps (**Figure 3.3**). In addition, we note that chromatin fragments generated by restriction enzymes in typical Hi-C protocols are significantly larger than mononucleosomes, increasing the number of crosslinking opportunities per fragment and thus

presumably restricting their diffusion and resultant ability to generate artefactual ligation products. We propose that this difference in fragment size/mobility accounts for the increased noise seen previously in Micro-C relative to standard Hi-C protocols. Further supporting this idea, we find that the improved protocol strongly reduces the incidence of artefactual ligation products between the nuclear genome and the mitochondrial genome, relative to the standard Micro-C protocol (**Figure 3.5A**). We note this is in general agreement with prior comparisons of in-solution versus both pellet and “in situ” Hi-C protocols (GAVRILOV *et al.* 2013; RAO *et al.* 2014; NAGANO *et al.* 2015a). Together, these considerations support the idea that the use of long crosslinkers and isolation of insoluble chromatin may be important to prevent mononucleosomes from freely diffusing prior to ligation and introducing noise into Micro-C measurements. Still, this does not rule out the additional possibility that in some cases our long crosslinkers capture nearby genomic loci for which the closest crosslinkable proteins are not in immediate physical proximity, and indeed both of these features may contribute to the improvement in data quality seen in Micro-C XL.

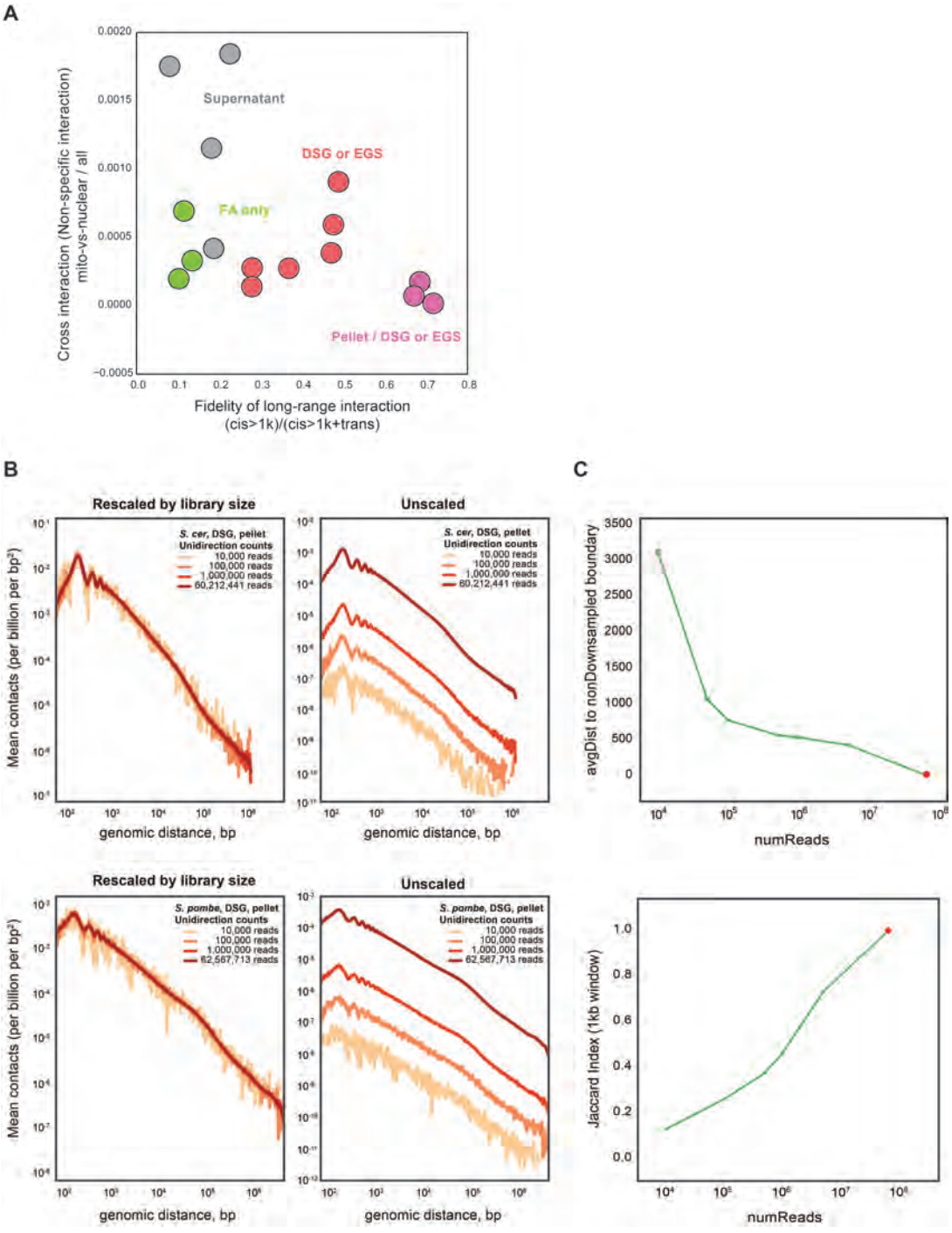


Figure 3.5. Considerations of sequencing depth for mammalian cells and subsampling Micro-C XL data.

(**A**) Effects of Micro-C protocols on artefactual interactions. For each Micro-C dataset generated for *S. cerevisiae* in this study, we calculated 1) the fraction of sequencing reads mapping to the mitochondrial genome, and 2) the ratio between those reads reporting on an interaction between two loci on the same chromosome (in cis, >1 kb), and reads reporting on an interaction between chromosomes (in trans). We restrict cis reads to those >1kb to avoid any influence of molecular byproducs on the cis/total ratio. Here, these two values are scatterplotted against one another for all Micro-C datasets. Note that supernatant libraries exhibit a greater frequency of mitochondrial reads relative to other Micro-C libraries, and that pellet libraries exhibit a strong depletion of trans interactions. (**B**) Plots of interaction vs. distance are extremely robust to downsampling of sequencing data. Left plots show normalized density of interactions per squared base pair (y axis, normalized to total number of reads) vs. distance (x axis) for data downsampled to 10,000, 100,000, or 1,000,000 reads, or for the entire dataset. Right panels show the four curves separately, without normalization to sequence depth. In all cases, Micro-C XL reads (DSG pellet) were downsampled (after removing PCR duplicates) to the indicated number of reads. (**C**) (Top panel) Micro-C XL reads (*S. cerevisiae*, DSG pellet) were downsampled as indicated (x axis), and boundaries between CIDs were called as previously described (HSIEH *et al.* 2015). Y axis shows distance between the boundary location called in the downsampled dataset and the nearest boundary called from the full dataset. Curves show average over ten subsampling analyses, and squares at each value of reads indicate the small standard deviation of these replicas. Note that even with only 100,000 sequencing reads, chromatin boundaries are identified to within 1 kb. The red circle represents the full dataset, where the average distance is zero. (Bottom panel) Downsampling results in monotonic loss of boundary information. Here, y axis shows the average Jacard index of CID boundaries called from the full dataset versus the downsampled dataset with the number of reads indicated on the x axis. Curves again show the average of ten subsampling analyses, and the red circle represents the full dataset, where the Jaccard index is one. We note that the performance of this particular boundary-calling method does not represent a fundamental limit on the recovery of domain boundaries from sparse datasets, which represents a possible topic for future computational methods.

Chromosome structure

Validated via comparisons to prior data, our method provides insight into yeast genome folding at all length scales of interest. At larger scales, the Rabl configuration of chromosomes is seen as clustering of centromeres, and interactions between the telomeres of chromosome arms of similar length. Centromeric chromatin also shows a characteristic “X” shape resulting from the two arms of the chromosome both statistically leading away from the centromere together for some ~20 kb, with centromeres otherwise being relatively isolated from chromosome arms. At higher resolution, genes in both budding and fission yeast are organized into chromosomally-interacting domains (CIDs), typically spanning 1-5 genes, that are in some ways similar to the “topologically-associating domains” described in a multitude of other model organisms. Boundaries between CIDs occur at active promoters, highly-expressed genes, and tRNA genes (**Figure 3.4G**). In both budding and fission yeast, genomic regions surrounding cohesin-associated loci are relatively insulated from physically interacting with one another (**Figure 3.4K**). However, this insulation is stronger and persists over greater genomic distances in *S. pombe* relative to *S. cerevisiae*, pointing towards important differences in the role of cohesin, potentially in a cell-cycle dependent fashion, between the species. Taken together, these analyses highlight the ability of Micro-C XL to assay chromosome folding across all scales, as well as its broad applicability and future utility in comparative genomics.

Applications to other biological systems

We finally turn to considerations of the sequencing depth required for applications of Micro-C XL in organisms with larger genomes. As reviewed in (LAJOIE *et al.* 2015) the fundamental genomic resolution of a chromosome capture dataset is set by the frequency at which the genome is fragmented prior to capture of physical interactions by ligation; beyond this lower bound to resolution, the effective genomic resolution is further influenced by sequencing depth and library complexity. The proportion of molecular byproducts in a library additionally influences the amount of sequencing required to achieve a given coverage per fragment. Given that Micro-C XL does not display a preponderance of molecular byproducts (**Figure 3.5A**), the sequencing depth required to achieve a given genomic resolution should be similar to a Hi-C protocol. Nevertheless, Micro-C XL has the capacity to analyze chromatin interactions at genomic distances smaller than currently available Hi-C protocols. Indeed, the highest-resolution studies performed to date in mammals (RAO *et al.* 2014) utilize restriction enzymes with 4 bp target sequences to yield average fragment lengths of ~256 bp, although due to the heterogeneous distribution of restriction sites across the genome lower-resolution (~1 or 5 kb) binning approaches must be used in analysis of such datasets. In comparison to any individual 4-cutter, MNase digestion of chromatin to mononucleosomes results in at most ~75% more genomic fragments (depending on the nucleosome repeat length in the

tissue of interest), which in turn increases the fundamental genomic *resolution* of Micro-C by substantially more than ~1.75-fold, thanks to the more even spacing of the resulting fragments.

In addition, beyond binning Micro-C data to mimic lower-resolution Hi-C, it is important to note that a wide variety of biological questions can be addressed – at high resolution – by Micro-C at much lower sequencing depth. First of all, the strength of the Micro-C protocol is its ability to interrogate chromatin fiber structure at ~150-1000 bp resolution – there is little reason to carry out Micro-C to investigate >1 MB chromatin domains (LIEBERMAN-AIDEN *et al.* 2009). A key measure in this regard (IMAKAEV *et al.* 2015) is the decay of interaction frequency with increasing distances (see, e.g., **Figure 3.2C** or **Figure 3.3C**), which is an averaged measure across the entire genome and is thus extremely robust to undersequencing (**Figures 3.5B-C**). We anticipate that very low coverage (below 1-2 million reads) Micro-C XL maps in mammals will thus allow robust comparison of average chromatin fiber folding for, say, Polycomb-repressed genes, or for exons vs. introns, etc.

In addition to using such computational averaging methods to make use of multiple instances of any given annotation, the molecular complexity of the sequencing library can also be experimentally reduced. This is commonly done in sequence-capture RNA-Seq protocols in cancer exome studies, and more recently such methods have been applied to 3C methods based either on protein capture (eg, Pol2 IP) or on capture of specific “bait” sequences (DENKER AND DE

LAAT 2016). The reduction in read depth required for such methods naturally depends on the distribution and abundance of the feature to be captured, but many proteins of interest – CTCF, cohesin, TFIIB, and others – are sparsely distributed enough to enable >100-fold reductions in the sequencing depth required for high-resolution Hi-C studies. These and other considerations (LAJOIE *et al.* 2015; DENKER AND DE LAAT 2016) must be a part of any experimental design for a 3C-based study.

Conclusion

Here, we describe a modified protocol for genome-wide analysis of 3D chromatin structure that captures aspects of chromosome folding at all scales from mononucleosome resolution up to interactions between different chromosomes. This protocol, Micro-C XL, should find broad utility in a multitude of biological systems.

Materials and Methods

Yeast strains and culture conditions

All experiments reported here were carried out with either *S. cerevisiae* strain BY4741 or *S. pombe* strain 972 h-. BY4741 cultures were grown in YPD media at 30°C, while *S. pombe* cells were grown at 30°C in “Compromise Media” (TSANKOV *et al.* 2010), consisting of Yeast extract (1.5%), Peptone (1%),

Dextrose (2%), SC Amino Acid mix (Sunrise Science) 2 g/L, Adenine 100 mg/L, Tryptophan 100 mg/L, and Uracil 100 mg/L.

Fixation conditions

Midlog yeast cultures were crosslinked with either 1% or 3% final concentration of formaldehyde (Sigma) for 15 minutes at 30°C, then quenched with 125 mM glycine for 5 min at room temperature. Yeast were then spheroplasted as previously described (YUAN *et al.* 2005; TSANKOV *et al.* 2010). For DSG and EGS (Thermo Fisher Scientific) crosslinking studies, spheroplasts were resuspended in a 3 mM final concentration of the crosslinker of interest in PBS, and crosslinked for 40 min at 30°C, then quenched with 125 mM glycine for 5 min at room temperature. Note that we use the protocol “Micro-C XL” for DSG and EGS-based protocols interchangeably, as the data from these protocols are nearly indistinguishable.

Separation of soluble and insoluble chromatin

Crosslinked chromatin was digested with micrococcal nuclease (MNase, Worthington) to yield > 95% mononucleosomes. After inhibition of MNase with 2 mM EGTA at 65°C, fragmented lysate was in some cases used directly for the standard Micro-C protocol, or in some experiments was separated into supernatant and pellet portions. Here, MNase-digested lysate was spun at

16,000 g for 5 minutes, and Micro-C library construction was separately performed on supernatant or on the pellet fraction.

Micro-C protocol

Briefly, the termini of nucleosomal DNA were dephosphorylated by Shrimp Alkaline Phosphatase and then subjected to T4 DNA polymerase for end repairing and biotin labeling by supplementing with biotin-dCTP, biotin-dATP, dTTP, and dGTP. Crosslinked chromatin was diluted to 10 ml and treated with T4 DNA ligase. After heat inactivation, chromatin was concentrated to 250 µl in an Amicon 30k spin column and treated with 100 U exonuclease III for 5 min to eliminate biotinylated ends of unligated DNA. Proteinase K was then added and incubated for 65°C overnight. DNA was purified by PCI extraction and ethanol precipitation, treated with RNase A, and ~250–350 bp DNA was gel-purified. Purified DNA was treated with End-it, subject to A-tailing with Exo-Klenow, and ligated to Illumina adapters. Adapter-ligated DNA was purified with streptavidin beads to isolate ligated Micro-C products away from undigested dinucleosomal DNA. Streptavidin beads were then subject to ~10–12 cycles of PCR using Illumina paired-end primers. Amplified library was purified and subject to Illumina NextSeq paired end sequencing.

Computational analysis of Micro-C interactions

Micro-C data was mapped to the sacCer3 genome using Bowtie 2.1.0 as described in (IMAKAEV *et al.* 2012) using the hiclib library for python, publicly available at https://bitbucket.org/mirnylab/hiclib, with virtual 100bp fragments. Raw interaction maps were plotted as described previously (HSIEH *et al.* 2015). To obtain corrected contact maps, genomic coverage was calculated by summing the total number of interactions per bin. Low coverage bins were then excluded from further analysis using a MAD-max (maximum allowed median absolute deviation) filter on genomic coverage, set to 9 median absolute deviations. Following this filtering, stand-alone bins were removed (ie. regions where both neighboring bins did not pass filters), and the resulting maps were then iteratively corrected to equalize genomic coverage (IMAKAEV *et al.* 2012). Observed/expected contact maps were obtained by additionally dividing out the dependence on genomic distance, calculated empirically as the mean number of contacts at each genomic separation, using a sliding window with linearly increasing size, as previously described (NAUMOVA *et al.* 2013). Log-log plots of contact probability *P(s)* (also termed interaction frequency) versus distance were calculated using log-spaced bins with a constant step size. For average plots around genomic features, gene positions and orientations, centromere positions, and tRNA positions were obtained from the SGD (http://www.yeastgenome.org/).

CHAPTER IV

High-Resolution Chromatin Dynamics during a Yeast Stress Response

Abstract

Covalent histone modifications are highly conserved and play multiple roles in eukaryotic transcription regulation. Here, we mapped 26 histone modifications genome-wide in exponentially growing yeast and during a dramatic transcriptional reprogramming – the response to diamide stress. We extend prior studies showing that steady-state histone modification patterns reflect genomic processes, especially transcription, and display limited combinatorial complexity. Interestingly, during the stress response we document a modest increase in the combinatorial complexity of histone modification space, resulting from roughly 3% of all nucleosomes transiently populating rare histone modification states. Most of these rare histone states result from differences in the kinetics of histone modification that transiently uncouple highly correlated marks, with slow histone methylation changes often lagging the more rapid acetylation changes. Explicit analysis of modification dynamics uncovers ordered sequences of events in gene activation and repression. Together, our results provide a comprehensive view of chromatin dynamics during a massive transcriptional upheaval.

Introduction

All genomic transactions in eukaryotes take place in the context of a chromatin template (KORNBERG AND LORCH 1999). Chromatin plays key regulatory roles in control of transcription and other processes, and a great deal of highly-conserved cellular machinery is devoted to manipulation of nucleosome positioning (JIANG AND PUGH 2009; HUGHES AND RANDO 2014), histone subunit composition (HENIKOFF AND AHMAD 2005), and covalent modification states (SUGANUMA AND WORKMAN 2008). Histone modifications play key roles in transcriptional control, cell state inheritance, and many other processes. Genome-wide maps of histone modifications exist for a variety of organisms, and have been used for identifying regulatory and functional elements of the genome (GUTTMAN *et al.* 2009; HON *et al.* 2009; ERNST *et al.* 2011).

Two outstanding questions in histone modification biology are raised by such genome-wide maps. First, histone modifications often occur at thousands of genomic locations (e.g., at every active transcription start site), yet typically have functional importance for transcription at a small subset of marked genes under standard growth conditions (LENSTRA *et al.* 2011; WEINER *et al.* 2012). This raises the question of how a gene's context – local sequence context and/or other histone modifications – impacts the functional readout of a given histone modification. The second question is why such a plethora of histone modifications are used by the cell – over 100 histone modifications have been identified, yet histone modifications co-occur in large, tightly correlated groups, and exhibit little combinatorial complexity (RANDO 2012).

Both of these observations – that histone modifications often occur at genes where they serve no apparent function, and that histone modifications co-occur – are at least partially the consequence of biological feedback. In other words, because transcript levels are buffered by feedback mechanisms, many of them are restored to wild type levels in deletion mutants. Similarly, histone modifications often co-occur as a result of histone modification "crosstalk", in which the enzyme that deposits mark B preferentially acts on A-marked nucleosomes (SUGANUMA AND WORKMAN 2008). Histone modification networks thus include many feedforward and feedback loops of varying degrees of complexity. One way to uncover mechanisms of homeostasis is to perturb a network and study the time-evolution of as many individual nodes in the network as possible – such observations can potentially distinguish direct effects from slower indirect effects.

Functional genetic studies confirm the value of extending steady-state studies to a dynamic context. Time course analyses of transcriptional response to perturbations have previously uncovered unanticipated roles for chromatin-related mutants – a multitude of single gene studies (see, e.g.,(KORBER *et al.* 2006)), as well as genome-scale studies (WEINER *et al.* 2012), have shown that chromatin regulators are more important during changes in transcription than they are for steady-state transcription.

These considerations lead us to further explore the effects of transcriptional reprogramming on histone modification dynamics. We used ChIP-

seq to systematically map dynamic changes of 26 histone modifications in response to a stress signal in yeast (**Figure 4.1A**). Our data recover known aspects of the steady-state histone modification landscape, and show that relationships between histone modifications and transcription are maintained during the stress response. Most interestingly, during the stress response roughly 3% of all nucleosomes occupy unusual regions of histone modification space that are unoccupied in steady state. Inspection of these nucleosomes identifies differences in the kinetics of different histone modifications, and reveals multiple stages of the chromatin response to transcriptional changes.

Results

Genome-wide patterns of covalent histone modifications

We focus here on the yeast response to the sulfhydryl reducing agent diamide, which involves rapid and massive transcriptional reprograming of both the common stress response genes and ~200 genes involved in cell wall protection and redox homeostasis (GASCH *et al.* 2000). Overall 19% of all mRNAs change expression during this response.

Using MNase-ChIP-seq (LIU *et al.* 2005; RADMAN-LIVAJA *et al.* 2011a) we mapped 26 histone modifications at mononucleosome resolution at varying times (t=0, 4, 8, 15, 30, and 60 minutes) after diamide addition (**Figure 4.1).** Easily apparent in the resulting genomic tracks (**Figure 4.1B**) are many well-described features of yeast chromatin, including: 1) Generally well-positioned nucleosomes,

2) nucleosome depletion at promoters, 3) H3K4me3 and acetylation enrichment at 5’ ends of genes, 4) H3K36me3 covering mid- and 3’ coding regions, and many more steady-state chromatin hallmarks. Zooming in on the dynamics during diamide response (**Figure 4.1C)** demonstrates typical behavior for the highly-induced gene *GLK1* with H3K4me3, H3K36me3, and other transcription-correlated marks increasing over the gene body over time. Conversely, H4K16ac decreases over *GLK1*, presumably as a result of increased histone turnover during transcriptional induction.

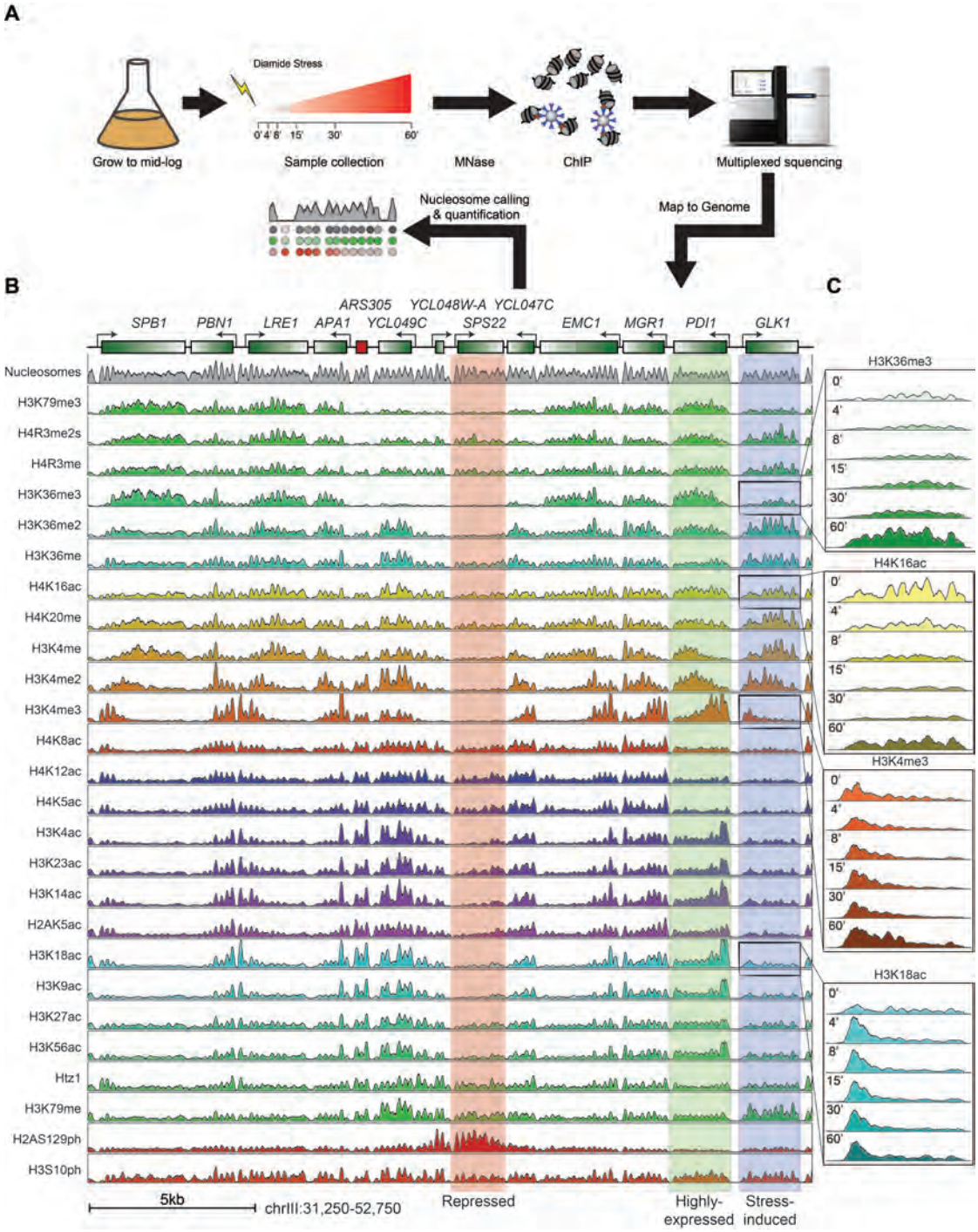


Figure 4.1. Epigenomic landscape of a yeast stress response.

(A) Experimental outline. Yeast were subject to 1.5 mM diamide stress, and cultures were harvested for MNase-ChIP-seq mapping at the indicated time points. **(B)** Steady-

state histone modification landscape for budding yeast. ChIP-Seq signal for 26 histone modifications and nucleosome mapping data. Top panel shows genomic annotations. **(C)** Histone modification dynamics over *GLK1*, a typical stress-induced gene. Data is shown for six time points following diamide stress.

Steady-state histone modifications follow stereotypical patterns

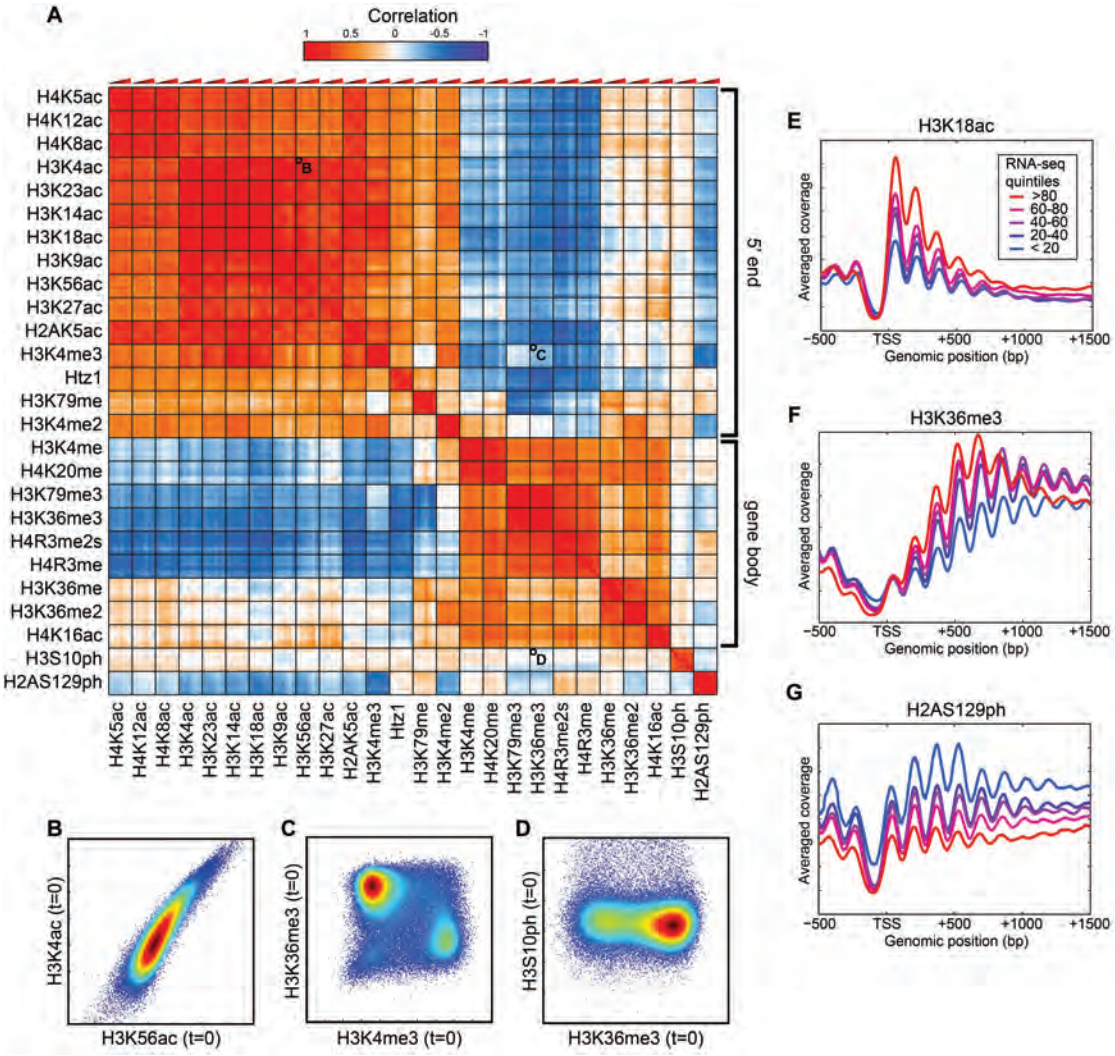
We start by refining models of the chromatin landscape. Using the nucleosome mapping data to call 66,360 nucleosomes, we calculated the occupancy of each nucleosome, as well as the level of the 26 mapped modifications, normalized to nucleosome occupancy.

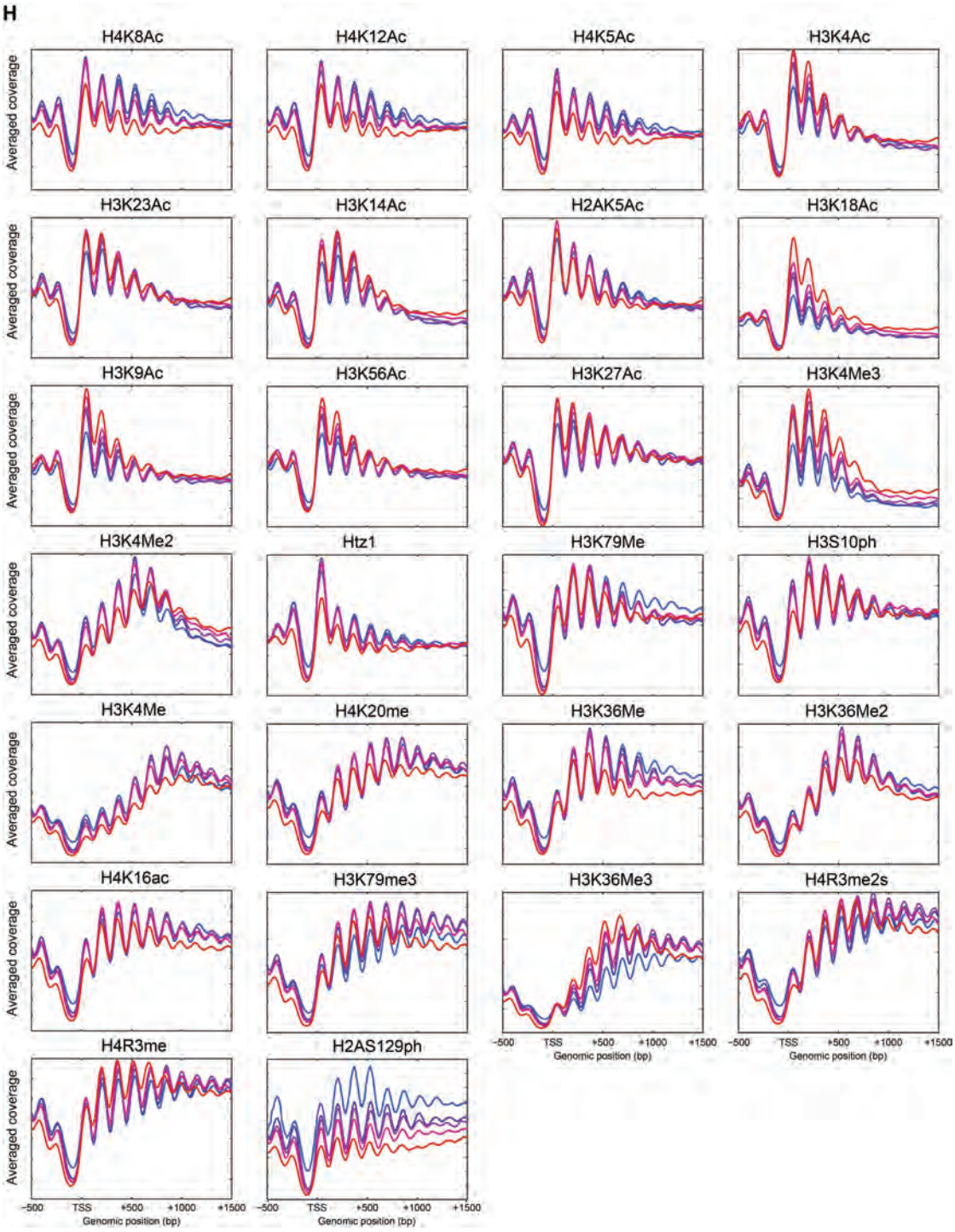
To explore the relationships between histone modifications, we calculated the correlation between the levels of all modifications, at all time points, across all nucleosomes (**Figures 4.2A-D**). The 156X156 correlation matrix (**Figure 4.2A)** shows a strong concordance between all 6 time points for each modification – each 6X6 box on the diagonal is bright red – indicating that the global genomic landscape of any given modification is not drastically altered by diamide stress. The 26 marks studied here show relatively few basic types of genomic modification patterns. This is consistent with previous observations in yeast (LIU *et al.* 2005), flies (FILION *et al.* 2010), and mammals (ERNST *et al.* 2011) of low combinatorial complexity among histone modifications. The majority of histone modifications are found in two large groups (**Figure 4.2A**). The first group includes H3K4me3 and lysine acetylation marks that occur at the 5' ends of coding regions and that scale with transcription rate (**Figure 4.2E** and **H**). The

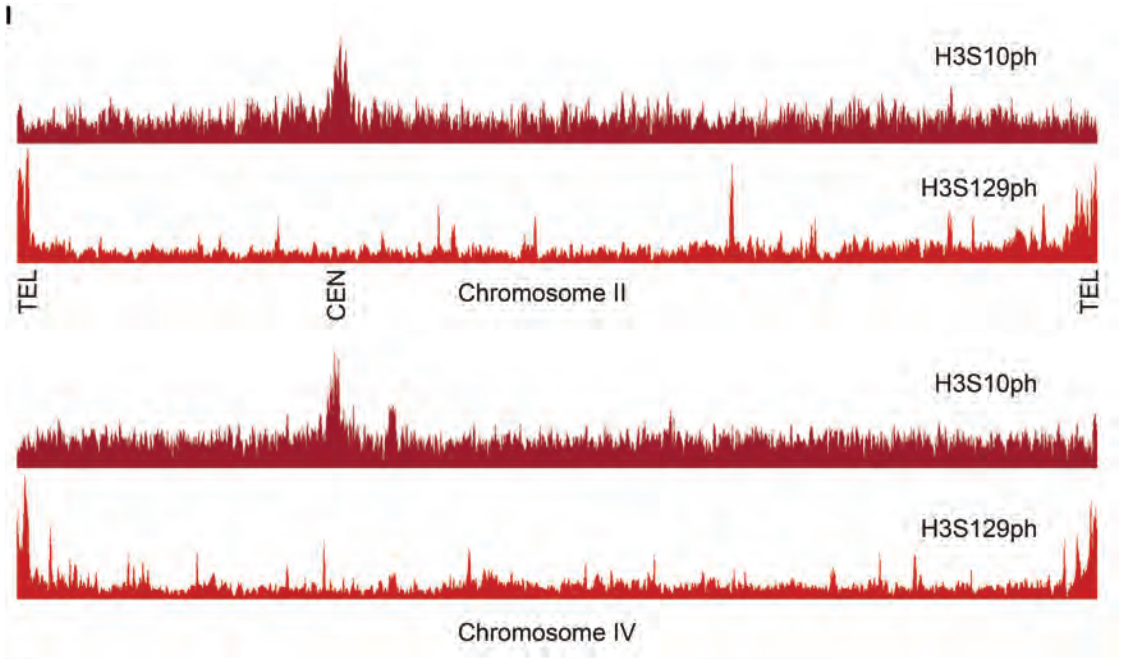
second (albeit less coherent) group in **Figure 4.2A** is of modifications occurring over middle and 3' ends of coding regions, such as H3K36me3 (**Figure 4.2F**). In addition, several modifications exhibit localization patterns related to gene structure that are somewhat distinct from the two main groups, including H4K16ac, the H2A variant Htz1 (H2A.Z), and various mono- and di-methylation marks (see below).

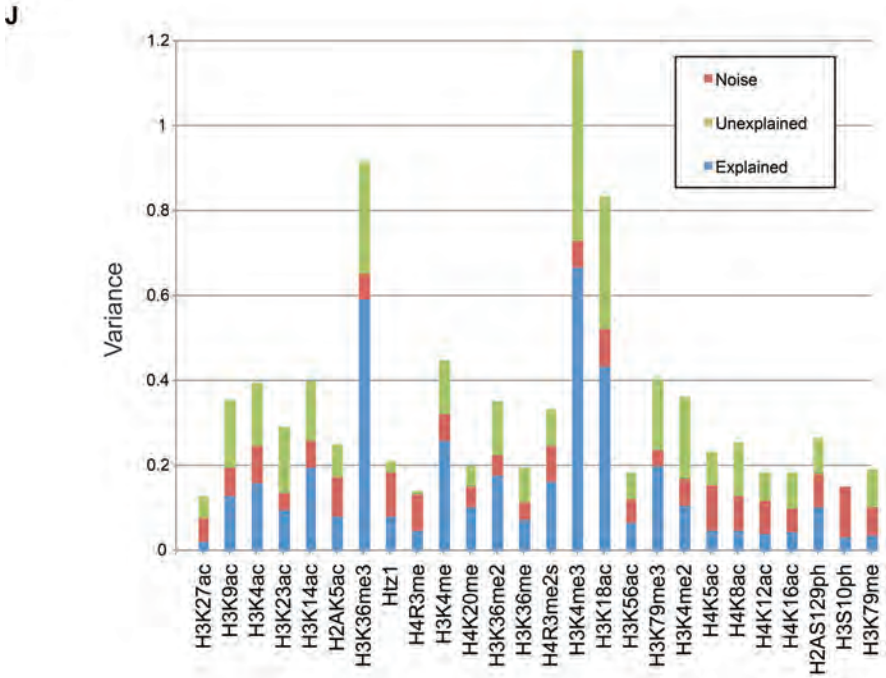
Finally, in addition to marks with localization patterns related to gene structure, the two phosphorylations stood out as unusual (**Figures 4.2A** and **G**). The localization pattern of H3S10ph is dominated by a ~20 kb pericentric domain (**Figure 4.2I**) deposited every M phase (CROSIO *et al.* 2002), a signal which overwhelms the minor coding region signal of anticorrelation with histone turnover (WEINER *et al.* 2012). The other unusual mark in this dataset is the DNA damage-related H2AS129ph modification (often referred to as γ-H2AX), which occurs over subtelomeric regions and actively repressed genes (SZILARD *et al.* 2010) (**Figures 4.1B**, **4.2G**, and **4.2I**). Analysis of the chromatin packaging state of rDNA genes (**Figure 4.2K**) shows similar correlation structure to that observed over the Pol2-transcribed regions of the genome.

Overall, our results recapitulate essentially all known aspects of the steady-state landscape of histone modifications in yeast, showing the quality of the dataset.


A
Correlation
1
0.5
0
-0.5
-1
H4K5ac
H4K12ac
H4K8ac
H3K4ac
H3K23ac
H3K14ac
H3K18ac
H3K9ac
H3K56ac
H3K27ac
H2AK5ac
H3K4me3
Htz1
H3K79me
H3K4me2
H3K4me
H4K20me
H3K79me3
H3K36me3
H4R3me2s
H4R3me
H3K36me
H3K36me2
H4K16ac
H3S10ph
H2AS129ph
5' end
gene body
B
C
D
E
H3K18ac
RNA-seq quintiles
>80
60-80
40-60
20-40
< 20
Averaged coverage
−500
TSS
+500
+1000
+1500
Genomic position (bp)
F
H3K36me3
G
H2AS129ph
H3K4ac (t=0)
H3K56ac (t=0)
H3K36me3 (t=0)
H3K4me3 (t=0)
H3S10ph (t=0)
H3K36me3 (t=0)


H
H4K8Ac
H4K12Ac
H4K5Ac
H3K4Ac
H3K23Ac
H3K14Ac
H2AK5Ac
H3K18Ac
H3K9Ac
H3K56Ac
H3K27Ac
H3K4Me3
H3K4Me2
Htz1
H3K79Me
H3S10ph
H3K4Me
H4K20me
H3K36Me
H3K36Me2
H4K16ac
H3K79me3
H3K36Me3
H4R3me2s
H4R3me
H2AS129ph
Averaged coverage
−500 TSS +500 +1000 +1500
Genomic position (bp)


I
H3S10ph
H3S129ph
TEL
CEN
Chromosome II
TEL
H3S10ph
H3S129ph
Chromosome IV


J
Variance
1.2
1
0.8
0.6
0.4
0.2
0
Noise
Unexplained
Explained
H3K27ac
H3K9ac
H3K4ac
H3K23ac
H3K14ac
H2AK5ac
H3K36me3
Htz1
H4R3me
H3K4me
H4K20me
H3K36me2
H3K36me
H4R3me2s
H3K4me3
H3K18ac
H3K56ac
H3K79me3
H3K4me2
H4K5ac
H4K8ac
H4K12ac
H4K16ac
H2AS129ph
H3S10ph
H3K79me

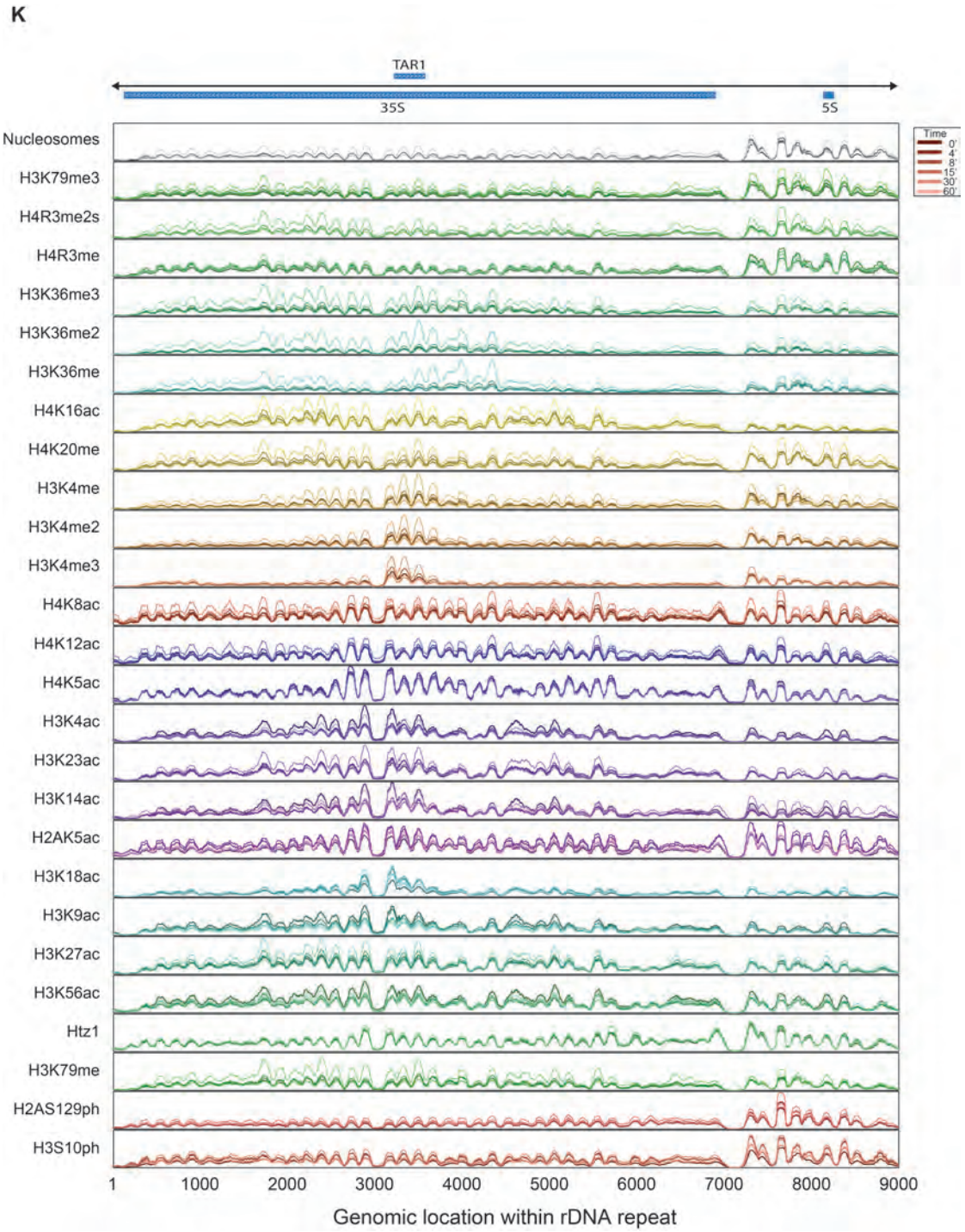


Figure 4.2. Characterization of histone modification patterns during midlog growth.

(A) Correlation matrix for 26 histone modifications. For each modification, 6 time points are arranged from t=0 to t=60 from left to right. **(B-D)** Scatterplots for strongly correlated (**B**), uncorrelated (**C**), and anticorrelated (**D)** pairs of modifications. Each scatterplot compares levels of the two modifications, normalized to nucleosome occupancy, for 66,360 individual nucleosomes in the yeast genome at t=0. Colors indicate density. **(E-G)** Metagene profiles for exemplary histone modifications. For each modification, data were aligned by the transcription start site (TSS) of annotated genes, grouped according to transcription rate (CHURCHMAN AND WEISSMAN 2011). **(H)** Metagene profiles at steady-state. For each modification, data were aligned by the transcription start site (TSS) of annotated open reading frames, and grouped according to transcription rate as in **Figures 4.2E-G**. **(I-J)** Features contributing to steady-state histone modification patterns. **(I)** Genome browser views for chromosomes II and IV, showing H3S10ph enriched around centromeres, and H2AS129ph enriched at telomeres, as indicated. **(J)** Contribution of experimental noise to the total variance in different histone modifications. For each modification dataset, we used the difference between adjacent diamide time points to estimate noise in the measurement, assuming that histone marks are relatively stable between adjacent time points. This naturally overestimates the noise in the dataset as yeast are changing their transcriptome and modifying the chromatin template in response to diamide stress. Nonetheless, this provides a reasonable measure of noise given that the features with the lowest noise here are generally associated with transcription-related marks (H3K4me3, etc.) which would be expected based on diamide stress to exhibit the most changes between time points. The higher noise estimated for marks such as H3S10ph and others thus likely reflects true measurement noise, either resulting from relatively nonspecific antibodies, or resulting from widespread marks with lower peak to trough values and thus lower “signal to noise”. **(K)** Chromatin landscape of rDNA repeats. The mapping of nucleosomes and 26 histone modifications in the 9.1kb rDNA repeat region. (top) Browser track showing transcripts from the repeat region. (bottom) Tracks showing enrichment of modifications along the repeat region. Line color saturation in each track ranges from dark (0’) to light (60’).

Steady-state histone modifications are mostly shaped by transcription and turnover

What are the major processes that shape the steady-state chromatin landscape? The first, detailed above, is the passage of RNA polymerase, which carries with it a large number of histone modifying enzymes as it traverses the genome (BURATOWSKI 2009). Second, genomic replication is pervasive and results in a dramatic but temporary restructuring of the chromatin template (GRUSS *et al.* 1993). Finally, histone modifications are altered by replication-independent histone turnover – newly synthesized histones are incorporated into the genome and replace old ones, thus removing old marks. New histones also carry a set of covalent modifications, some of which are deposited by enzymes that act specifically on free, but not nucleosomal, histones (DION *et al.* 2007; RUFIANGE *et al.* 2007; KAPLAN *et al.* 2008; DEAL AND HENIKOFF 2010). Beyond these pervasive processes, locus-specific processes can target specific chromosome positions, as for example observed at the pericentric domain of H3S10ph.

To quantify the extent to which the chromatin landscape is explained by these processes, we built a regression model that predicts the modification state of each nucleosome at midlog, based on the nucleosome position along the gene, its occupancy levels, the RNA polymerase level (in sense and anti-sense directions (CHURCHMAN AND WEISSMAN 2011)), the nucleosome's turnover rate (DION *et al.* 2007), its position relative to the centromere or telomeres, and its relative replication timing (RAGHURAMAN *et al.* 2001). This model explains 58% of

the overall variation in the 26-dimensional histone modification dataset, although the success of the model varies for different histone marks (**Figure 4.3A**). Some of the poorly explained modifications have higher inherent levels of measurement noise (**Figure 4.2J**), suggesting either a lower-quality antibody, or a relatively even genomic distribution of the histone mark in vivo. Note that the reported percentages are based on a relatively simple linear regression model, and represent a lower bound on predictive power.

To quantify the contribution of each feature to these predictions, we removed each feature in turn from the overall model, re-learned model parameters with remaining features, and determined the loss in variance explained (**Figure 4.3B**). These differences highlight the *unique* contribution of the removed feature to explaining a given process. These unique contributions do not necessarily sum to 100% of the signal, as many modifications are partially explained by several features (such as transcription and turnover, which are not purely uncorrelated with one another). In this case, removal of a single contributing feature will be partly compensated by other features in the re-learned model. Therefore, we describe the remaining fraction of 100% as synergistic interactions between features.

The most informative feature in our model was nucleosome position within the gene, consistent with the observation that many histone modifying enzymes are recruited to genes by either the initiation or elongation form of RNA Pol2 (BURATOWSKI 2009). The second most informative feature was transcription rate,

which predicted both well-characterized transcription-deposited marks, as well as marks that are anti-correlated with transcription (**Figure 4.3C**). Replication-independent turnover strongly predicts H3K56ac, a known mark of soluble histone proteins (TSUBOTA *et al.* 2007), as well as other marks of new histones such as H3K4ac and H3K9ac (GUILLEMETTE *et al.* 2011). Turnover was also predictive for slowly accumulating marks that are enriched in older nucleosomes – H3K79me3 and H4K16ac (**Figure 4.3C**). Chromosomal position was the best predictor of the pericentric H3S10ph and telomeric H2AS129ph marks (**Figures 4.3C** and **4.2I**), but unexpectedly also contributed to prediction of H3K79me levels. Replication timing explained little overall variance – as expected given that every locus in the genome is duplicated once per cell cycle – with its strongest explanatory power for the subtelomeric H2AS129ph mark.

While our analysis recapitulates many known features of chromatin, many additional connections are also documented. Most surprisingly, comparisons of sense and antisense transcription revealed a dichotomy among transcription-correlated marks between methylation and acetylation marks. Gene body marks were correlated mostly with sense transcription, while 5' marks appeared to read out total Pol2 transit in both directions (**Figures 4.3B-C**). This likely reflects rapid termination of inappropriate antisense transcripts (XU *et al.* 2009), which would prevent Pol2 from transitioning from initiation to elongation modification states.

Analysis of this predictive model shows subtle differences between highly correlated marks. For example, both H3K36me3 and H3K79me3 are

transcription-correlated gene body marks (**Figures 4.2A** and **4.2H**). However, their levels are explained to a different extent by transcription and turnover (**Figure 4.3C**), with H3K36 methylation mostly explained by genic position and sense transcription levels, while H3K79me3 levels are far more influenced by turnover rates. This likely reflects the fact that there is no known H3K79 demethylase, and thus this mark is presumably removed only by nucleosome eviction (DE VOS *et al.* 2011; RADMAN-LIVAJA *et al.* 2011b).

Histone modifications predict genomic processes

The fact that processes such as transcription and turnover are predictive of histone modifications allows the reverse – prediction of genomic transactions from chromatin data – to be used as a powerful experimental tool (GUTTMAN *et al.* 2009; HON *et al.* 2009; ERNST *et al.* 2011; GARBER *et al.* 2012). We thus asked whether the larger number of modifications mapped here identify more precise predictors for genomic processes. We applied sparse linear regression (**Methods**) to predict genomic features based on genome-wide modification data. The regression coefficients of this model confirm the expectation that the most predictive histone modifications are usually the ones that are most closely associated with the genomic process (**Figure 4.3D**). Thus, for example, H3K18ac, H3K4me3, and H3K36me3 are predictive of transcription, with positive regression weights, while H3K79me3 and H2AS129ph are also predictive, with negative weights. The best predictor of turnover rates is H3K79me3, which is

negatively correlated with turnover, while H3K18ac and H3K56ac provide positively-correlated predictors of turnover (**Figures 4.3D-F**).

Histone modifications are also predictive of replication timing. In particular, H3K56ac and H3K9ac – both marks associated with new nucleosomes – have opposite weights in this prediction. Higher H3K56ac levels are predictive of early replication times, while higher H3K9ac is predictive of later replication time. The connection between H3K56ac and early replication reflects both the length of time between a locus' replication time and M phase H3K56 deacetylation (CELIC *et al.* 2006; MAAS *et al.* 2006), as well as the high turnover characteristic of early origins (KAPLAN *et al.* 2008). The connection between H3K9ac and late replication is less clear – newly synthesized histone H3 is enriched for H3K9ac during S phase (ADKINS *et al.* 2007), but whether this mark is generated preferentially later during replication or is otherwise targeted to late-replicating domains is presently unknown.

Taken together, these analyses provide an expanded list of marks to be used for annotation of genomic features and processes.

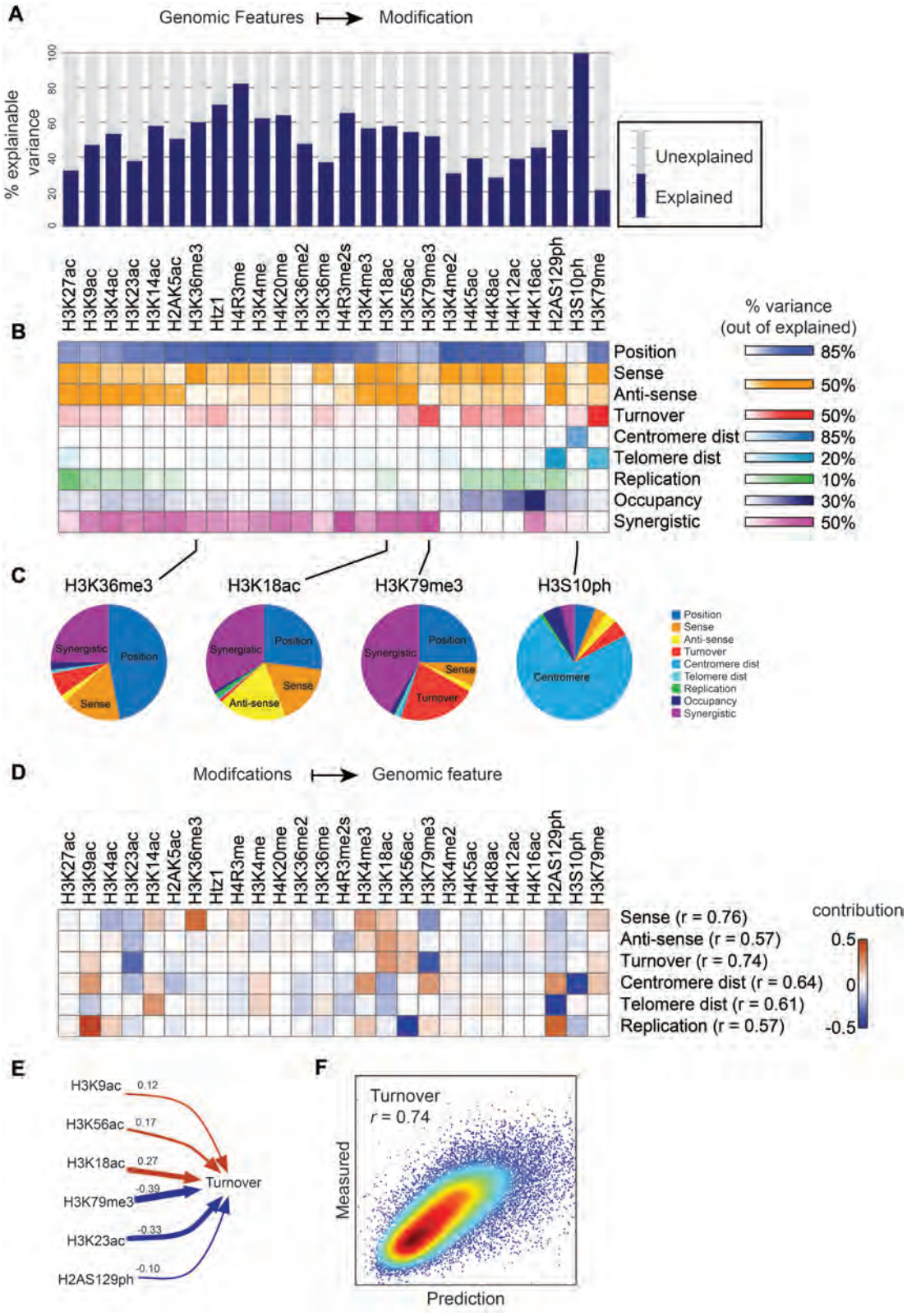

A
Genomic Features
Modification
% explainable variance
Unexplained
Explained
H3K27ac
H3K9ac
H3K4ac
H3K23ac
H3K14ac
H2AK5ac
H3K36me3
Htz1
H4R3me
H3K4me
H4K20me
H3K36me2
H3K36me
H4R3me2s
H3K4me3
H3K18ac
H3K56ac
H3K79me3
H3K4me2
H4K5ac
H4K8ac
H4K12ac
H4K16ac
H2AS129ph
H3S10ph
H3K79me
B
% variance
(out of explained)
Position 85%
Sense 50%
Anti-sense 50%
Turnover 50%
Centromere dist 85%
Telomere dist 20%
Replication 10%
Occupancy 30%
Synergistic 50%
C
H3K36me3
H3K18ac
H3K79me3
H3S10ph
Synergistic
Position
Sense
Anti-sense
Turnover
Centromere
Centromere dist
Telomere dist
Replication
Occupancy
D
Modifcations
Genomic feature
Sense (r = 0.76)
Anti-sense (r = 0.57)
Turnover (r = 0.74)
Centromere dist (r = 0.64)
Telomere dist (r = 0.61)
Replication (r = 0.57)
contribution
0.5
0
-0.5
E
H3K9ac 0.12
H3K56ac 0.17
H3K18ac 0.27
H3K79me3 -0.39
H3K23ac -0.33
H2AS129ph -0.10
Turnover
F
Turnover
r = 0.74
Measured
Prediction

Figure 4.3. Determinants of the steady-state modification landscape

(A) Predicting modification data from genomic features. A model incorporating genomic features (sense and antisense transcription (CHURCHMAN AND WEISSMAN 2011), nucleosome turnover rate (DION *et al.* 2007), distance from centromere and telomere, replication timing (RAGHURAMAN *et al.* 2001), and nucleosome position) predicts genomic patterns of all 26 histone marks. Plot shows the percent of signal explained per histone modification (see **Figure 4.2J**). **(B)** Contribution of genomic processes to explanatory power of the model. Heatmap shows the percentage of explained signal that is lost when a given process is removed from the model. Synergistic refers to remaining explained variance not lost upon removing any single feature. **(C)** Piecharts showing the variance explained by different aspects of the model for the indicated modifications. **(D)** Predicting genomic features from modification data. For each entry, the heatmap shows the sparse linear regression coefficient for the mark in question. **(E)** Turnover model parameters from (**D**) are shown here in numeric form. **(F)** Turnover model accurately captures turnover rates genome-wide. Model predictions (x axis) are scatterplotted against experimental turnover data (y axis).

Dynamics of individual histone modifications during transcriptional reprogramming

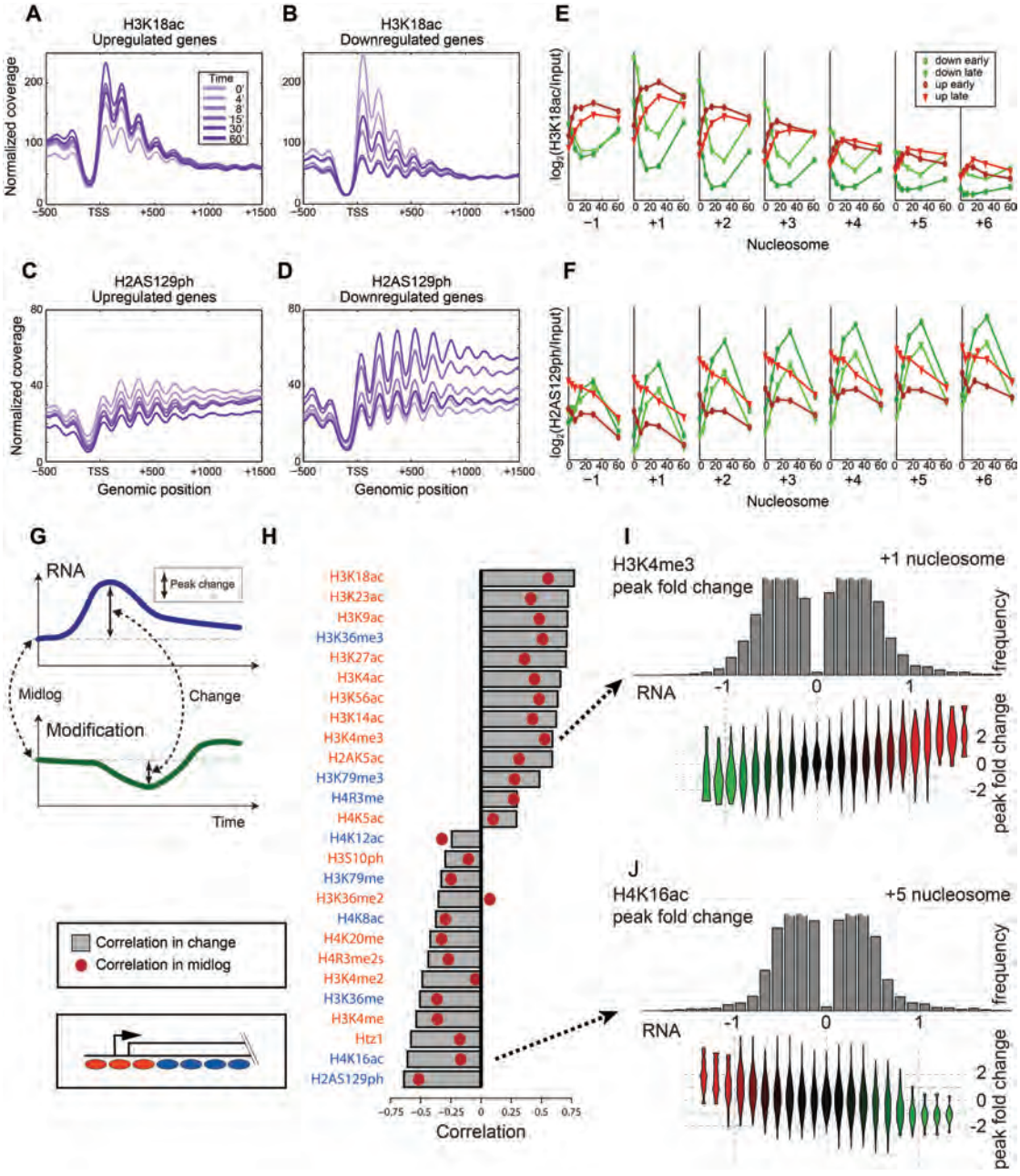
We next asked how individual histone modifications change during genome-wide transcriptional reprogramming (GASCH *et al.* 2000). We evaluated the nucleosome-specific change in each modification in terms of both the change relative to the t=0 level, and the extent to which the six time points show a consistent trajectory (**Methods**). At a 10% FDR, we find that many nucleosomes change in at least one modification (~60% of all nucleosomes), but substantially fewer show changes in several modifications, with ~7% changed in five or more marks (**Figures 4.4K-Q**). As expected, reprogrammed genes are significantly

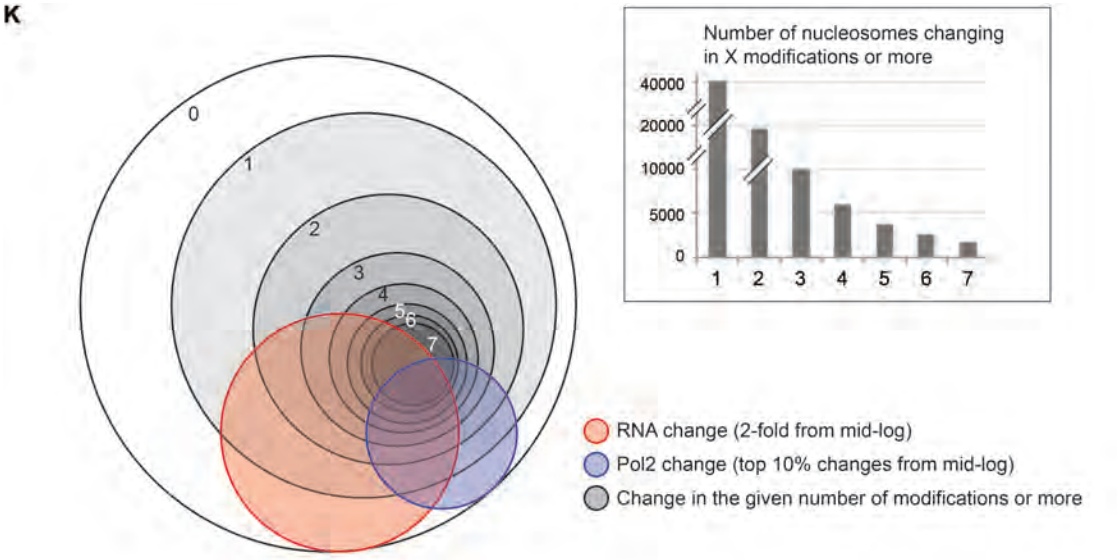
enriched with changed nucleosomes. Although these numbers reflect changes across a large fraction of the genome, the gross chromatin landscape features changed little during the stress response – see correlations for each modification across all six time points in **Figure 4.2A**.

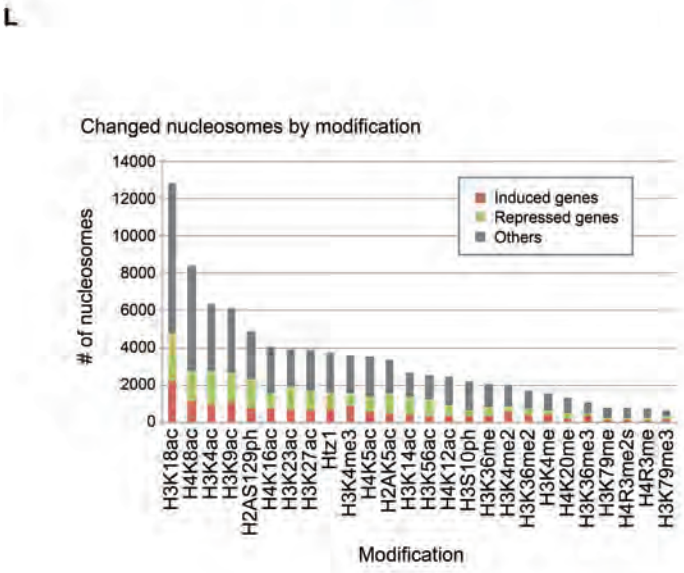
Next, we asked whether the relationship between histone modifications and transcription rate was altered by transcriptional reprograming. Steady state correlations observed between modifications and transcription might be universal to Pol2 passage itself, or alternatively could be linked to the specifics of the mid-log transcriptional program. Supporting the former model, we find that stress-induced transcriptional reprogramming generally maintains the mid-log relations between histone modifications and transcription rate. For example, levels of H3K18ac, a 5' mark correlated with transcription rate in mid-log (**Figure 4.2E**), increase at the 5' ends of activated genes and decrease over repressed genes (**Figures 4.4A-B**). Most other transcription-correlated marks have similar patterns. Similarly, the anti-correlation between H2AS129ph and transcription rate is also dynamic (**Figures 4.4C-D**), consistent with a previous study showing H2AS129ph being gained at *GAL* genes upon repression (SZILARD *et al.* 2010). In both cases, the peak of histone modification change coincided at t=30 with the peak change in mRNA and Pol2 levels (GASCH *et al.* 2000; KIM *et al.* 2010), before levels of all three fall at t=60 as yeast acclimate to the stressful environment (**Figures 4.4E-F**).

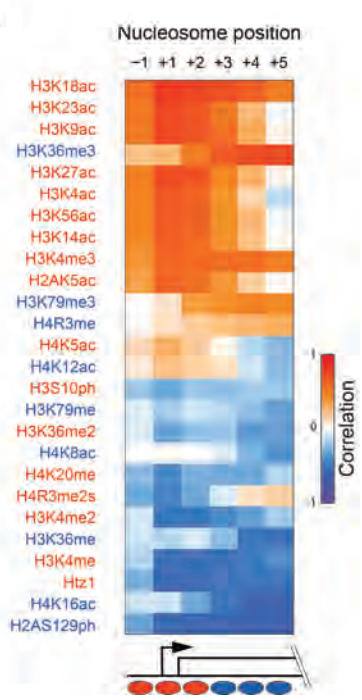
To visually compress the relationship between transcriptional dynamics and chromatin dynamics, for each histone modification and at each nucleosome position (+1, +2, etc.) we calculated the correlation between stress-induced changes in modification level, and changes in transcription (**Figures 4.4G-H** and **4.4M**). This analysis reveals the expected positive correlations between H3 acetylation states and transcription, and anti-correlation between H2AS129ph and transcription. Globally, we find strong concordance between the correlations of modifications with transcription rate at steady-state, and the correlation of the *changes* in modification levels with the transcriptional response (**Figures 4.4H-J**).

Notable exceptions to this general trend are two di-methylation states (H3K4me2 and H3K36me2), the histone variant Htz1 (H2A.Z), and H4K16ac, all of which are mostly uncorrelated with expression in steady state measurements but exhibit anti-correlation with transcriptional changes. These observations highlight the power of our approach to identify transient chromatin states – dimethylation states occur transiently during the accumulation of trimethylation as gene expression increases, and thus exhibit changes during reprogramming but are not captured at steady state. The transient changes in H4K16ac and Htz1, both of which have well-established links to histone turnover, likely reflect a transient phase of turnover during gene induction/repression.


A
H3K18ac
Upregulated genes
Normalized coverage
Time
0'
4'
8'
15'
30'
60'
−500
TSS
+500
+1000
+1500
B
H3K18ac
Downregulated genes
E
log2(H3K18ac/Input)
down early
down late
up early
up late
−1
+1
+2
+3
+4
+5
+6
Nucleosome
C
H2AS129ph
Upregulated genes
Genomic position
D
H2AS129ph
Downregulated genes
F
log2(H2AS129ph/Input)
G
RNA
Peak change
Midlog
Change
Modification
Time
H
H3K18ac
H3K23ac
H3K9ac
H3K36me3
H3K27ac
H3K4ac
H3K56ac
H3K14ac
H3K4me3
H2AK5ac
H3K79me3
H4R3me
H4K5ac
H4K12ac
H3S10ph
H3K79me
H3K36me2
H4K8ac
H4K20me
H4R3me2s
H3K4me2
H3K36me
H3K4me
Htz1
H4K16ac
H2AS129ph
−0.75 −0.5 −0.25 0 0.25 0.5 0.75
Correlation
Correlation in change
Correlation in midlog
I
H3K4me3
peak fold change
+1 nucleosome
frequency
RNA
peak fold change
J
H4K16ac
peak fold change
+5 nucleosome


K
Number of nucleosomes changing in X modifications or more
40000
20000
10000
5000
0
1 2 3 4 5 6 7
0
1
2
3
4
5
6
7
RNA change (2-fold from mid-log)
Pol2 change (top 10% changes from mid-log)
Change in the given number of modifications or more


L
Changed nucleosomes by modification
# of nucleosomes
14000
12000
10000
8000
6000
4000
2000
0
Induced genes
Repressed genes
Others
H3K18ac
H4K8ac
H3K4ac
H3K9ac
H2AS129ph
H4K16ac
H3K23ac
H3K27ac
Htz1
H3K4me3
H4K5ac
H2AK5ac
H3K14ac
H3K56ac
H4K12ac
H3S10ph
H3K36me
H3K4me2
H3K36me2
H3K4me
H4K20me
H3K36me3
H3K79me
H4R3me2s
H4R3me
H3K79me3
Modification


M
Nucleosome position
−1 +1 +2 +3 +4 +5
H3K18ac
H3K23ac
H3K9ac
H3K36me3
H3K27ac
H3K4ac
H3K56ac
H3K14ac
H3K4me3
H2AK5ac
H3K79me3
H4R3me
H4K5ac
H4K12ac
H3S10ph
H3K79me
H3K36me2
H4K8ac
H4K20me
H4R3me2s
H3K4me2
H3K36me
H3K4me
Htz1
H4K16ac
H2AS129ph
Correlation

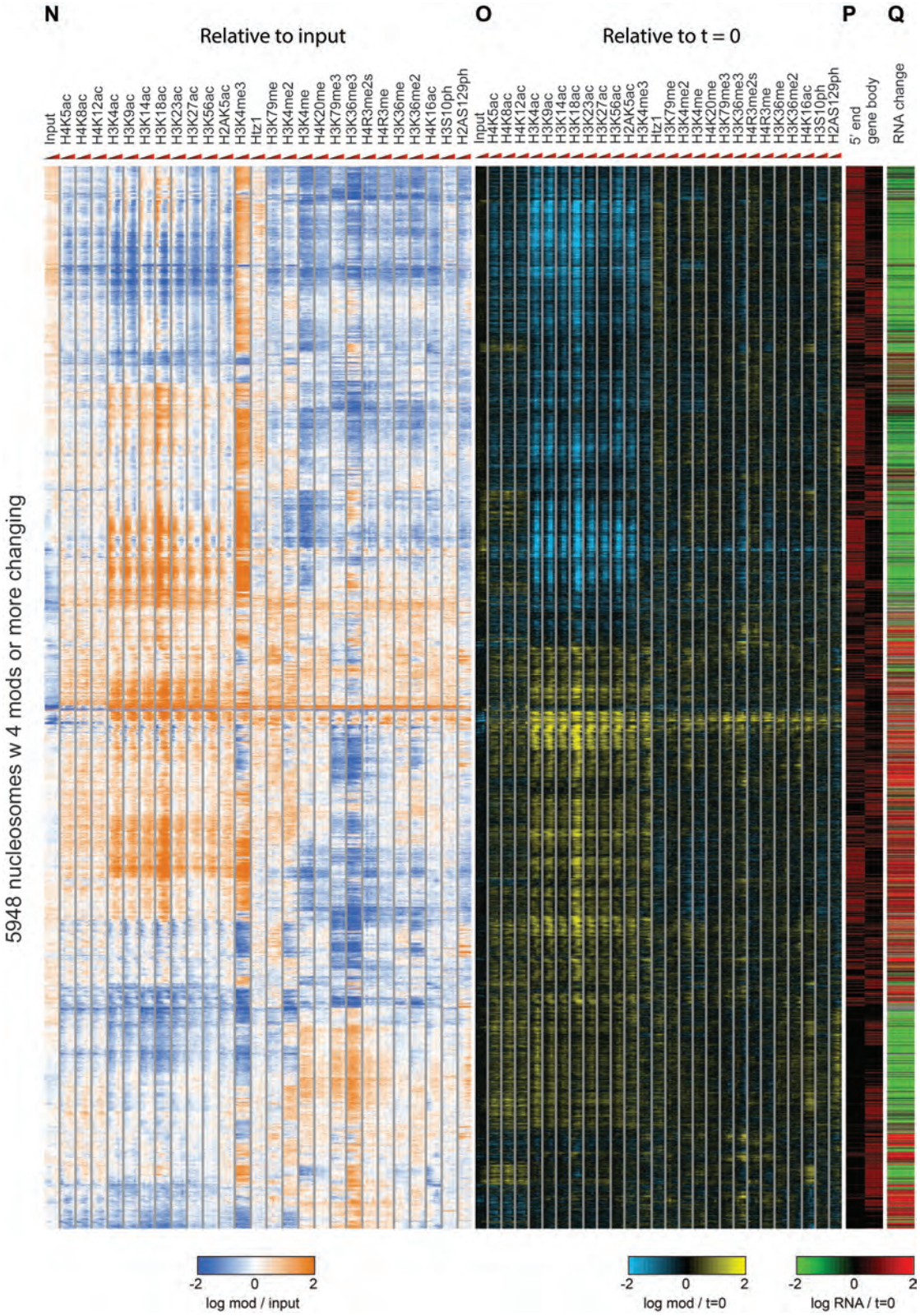

N
Relative to input
O
Relative to t = 0
P
Q
Input
H4K5ac
H4K8ac
H4K12ac
H3K4ac
H3K9ac
H3K14ac
H3K18ac
H3K23ac
H3K27ac
H3K56ac
H2AK5ac
H3K4me3
Htz1
H3K79me
H3K4me2
H3K4me
H4K20me
H3K79me3
H3K36me3
H4R3me2s
H4R3me
H3K36me
H3K36me2
H4K16ac
H3S10ph
H2AS129ph
5' end
gene body
RNA change
5948 nucleosomes w 4 mods or more changing
-2
0
2
log mod / input
log mod / t=0
log RNA / t=0

Figure 4.4. Dynamics of histone modifications during the stress response. **(A-B)** Metagenes showing levels of the transcription-correlated H3K18ac mark, averaged for upregulated (**A**) or downregulated (**B**) genes in response to diamide stress. **(C-D)** As in (**A-B**), for the repression-correlated H2AS129ph modification. **(E-F)** Dynamics of H3K18ac (**E**) or H2AS129ph (**F**) changes over time are shown averaged for various nucleosome positions along a gene body – the -1, +1, +2, etc. nucleosomes – as indicated. For each nucleosome, time course data for the modification in question are averaged for genes upregulated, or downregulated, relatively rapidly or slowly (**Methods**). **(G)** Schematic of approach to correlations between histone modification dynamics and transcriptional dynamics. **(H)** Correlations calculated as shown in (**G**), with red dots showing midlog correlations, and grey bars showing correlations between change in modification and change in transcription. **(I-J)** The correspondence between modification changes during diamide stress and transcription changes. In each case, a specific nucleosome location (+1, +5) as indicated. Top panel: histogram of the maximal change in the listed modification. Bottom panel: violin plots of changes in mRNA abundance for the genes carrying the nucleosomes in the bins above. **(L-M)** Stress-induced changes in histone modifications. **(K)** Venn diagram showing the overlap between nucleosomes on genes that changed expression or Pol2 levels (red and blue circle respectively) to nucleosomes that changed in at least 1, 2, …, 7 modifications (gray circles). Inset: histogram of the number of nucleosomes in each gray circle. **(L)** The number of nucleosomes that show significant movement in each of the modification. These numbers are broken according to location within induced, repressed genes, and other locations. **(M)** Correlation of change in expression to change of modification (as in **Figure 4.4H**) broken by nucleosome position. **(N-Q)** Patterns of stress-induced changes in histone modifications. Hierarchical clustering of histone modification patterns for 5948 nucleosomes with four or more changing marks (see **Figure 4.4K**). Each row is a nucleosome, and columns as follows. **(N)** Time course values of input levels (relative to median levels), and histone marks relative to input. **(O)** Time course values of input and histone marks relative to their levels at t=0. **(P)** Annotation of the nucleosome as 5’ or gene body. **(Q)** Maximum change in RNA for the associated gene during diamide response. Gray cell denote missing values. Although clustering was performed on the

values of (A) and (B), they form coherent clusters in terms of nucleosome position and direction of RNA change.

A modest increase in combinatorial complexity during the stress response

A key goal of this study was to determine whether tracing the time evolution of the histone modification network following a perturbation could uncover regulatory mechanisms (**Figure 4.5A**). We therefore sought to determine to what extent new combinatorial histone modification patterns appear during the stress response. We used Principal Component Analysis (PCA) to represent our 26-dimensional t=0 (steady-state) dataset, finding that three principal components could account for 76% of the variance in this dataset (**Figure 4.5B**). Interestingly, the variance in the dataset explained by these components decreased somewhat to ~67% during the early (8-15 min) response to diamide, before recovering nearly to baseline at the final time point of this response (**Figure 4.5B**). This increase in signal which is not explained by these three principal components indicates a transient increase in combinatorial complexity. Moreover, visualizing nucleosomes with the same two principal components for each time point revealed that nucleosomes transiently populate previously sparse regions of this 2D space early in the stress response (**Figure 4.5C**, black arrows).

To better understand such chromatin state transitions and how they are related to regulatory mechanisms, we further analyze 1) the shape of histone modification space, 2) the trajectories of nucleosomes through modification space (**Figure 4.5D**), and 3) kinetic distinctions between different histone marks.

A small number of unusual histone mark pairwise combinations

To identify changes in histone modification space in response to stress we asked what histone modification combinations are unique to, or at least enriched during, the early stages of the stress response. We estimated the density of the 26-dimensional space defined by nucleosome states at t=0, then sought nucleosomes that relocalize during stress to regions of this space that are sparsely-populated at t=0. Based on this 26D space, we identified ~2000 (3%) nucleosomes that moved to these low density regions during the stress response (**Methods**).

What novel histone modifications occur during the stress response? We considered 2-dimensional "slices" of this histone modification space for pairs of histone modifications (**Figure 4.5D**). In such space, we can distinguish between nucleosomes that move about inside the high-density region (e.g., **a** and **b** in **Figure 4.5D**) and ones that start inside the region and move outside during the response (e.g., **c** and **d** in **Figure 4.5D**). For example, 459 nucleosomes in the H3K4me3/H3K18ac space leave the high-density region out of 14,926 nucleosomes that change in this space (**Figure 4.5E**). Color coding of a nucleosome's location at t=0 allows the rough trajectory of the unusual nucleosomes at t=30 to be understood.

Analysis of all pairwise combinations identifies dramatic changes occurring for pairs of 1) H3K4me3 with various acetylation marks such as

H3K18ac or H3K14ac, 2) Htz1 with acetylation marks and gene body tri-methylation marks, and 3) H3K56ac with a range of marks. Although an average of 47 nucleosomes occupy rare regions of 2D space for each of the 325 pairwise modification combinations, the same nucleosomes are outliers in many separate 2D comparisons. Clustering of all rare nucleosomes in 26D space reveals a handful of behaviors that result in unusual modification combinations (**Figure 4.5I-M**), the most prevalent of which is the disconnect between H3K4me3 and histone acetylation marks.

While H3K4me3 and H3K18ac are normally extremely well-correlated (**Figures 4.2A** and **4.5E**), during diamide stress we find scores of nucleosomes carrying high levels of H3K4me3 but lacking H3K18ac, as well as the converse situation with highly-acetylated nucleosomes lacking H3K4me3 (**Figure 4.5E**). These nucleosomes are enriched at stress-repressed and -induced genes respectively.

A qualitatively distinct behavior from the H3K4me3/H3K18ac disconnect is seen for Htz1 and H3K56ac (**Figure 4.5F**). Although at t=0 these modifications are correlated, during the stress response we see two groups of nucleosomes that move into either the H3K56ac enriched / Htz1 depleted region (top left) or the opposite region (bottom right). Both groups of nucleosomes start with mild enrichment of both marks, meaning that during stress they gain one modification at the expense of the other. This behavior may result from a delay between H3/H4 replacement and Htz1 incorporation – Htz1 levels are low at promoters

with the highest H3/H4 turnover (GUILLEMETTE *et al.* 2005; DION *et al.* 2007) – as here an increase (for example) in H3K56ac indicates increased H3/H4 turnover that would also displace Htz1. In contrast, reduction in H3K56ac could speed up Htz1 accumulation by decreasing the ability of the SWR complex to carry out futile Htz1/H2A replacement cycles (WATANABE *et al.* 2013). We thus speculate that the transient disconnect between these marks results from a delay between stress-induced turnover and SWR recruitment, or vice versa.

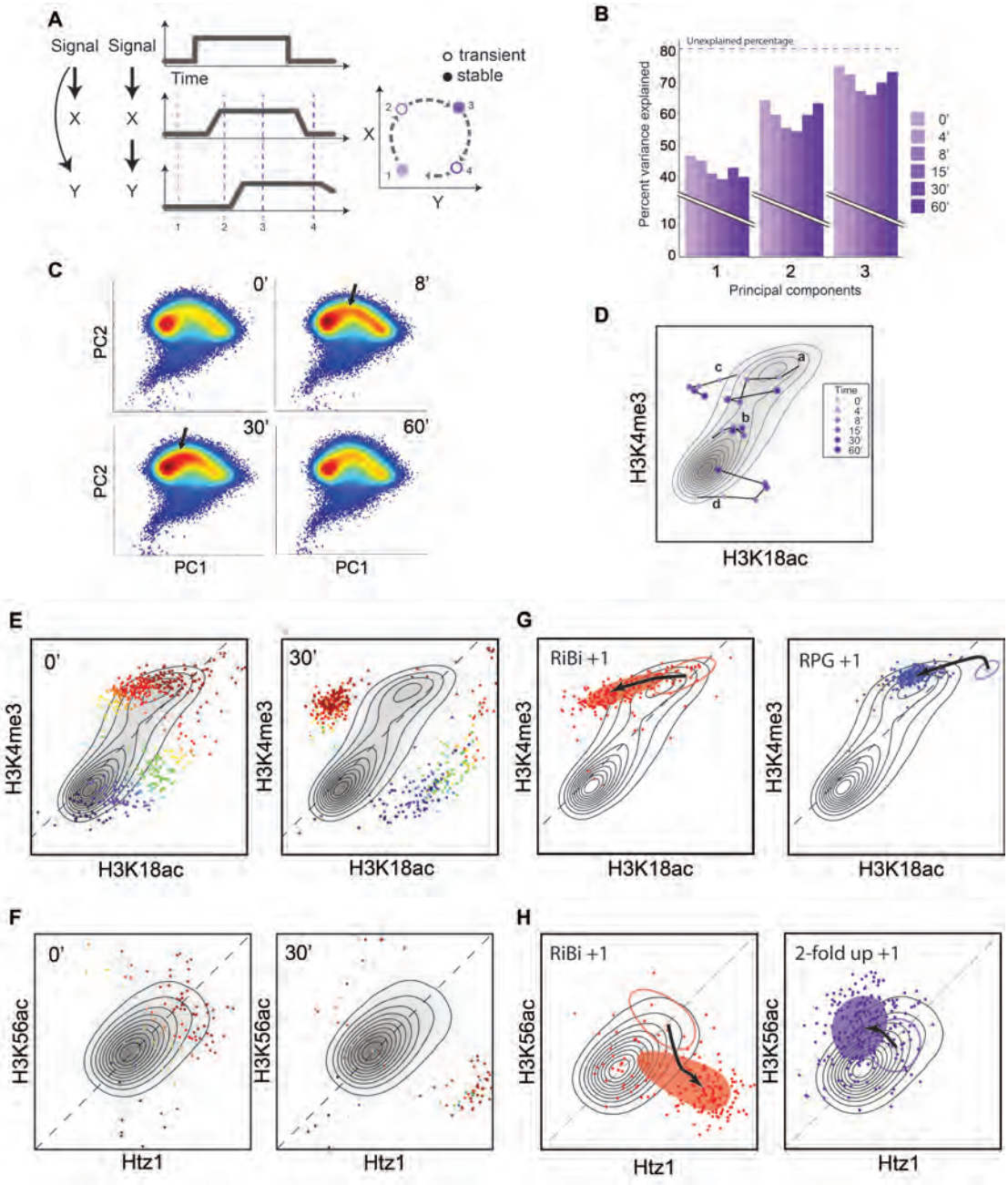
These results reveal, firstly, that ~97% of nucleosomes do not explore novel areas of histone modification space (for the 26 modifications profiled here, at our time resolution) even in response to a dramatic transcriptional perturbation in which 60% of all nucleosomes change levels of at least one modification. The remaining 3% of nucleosomes do transiently gain novel combinations of histone marks during the stress response, with 3-4 possible ways of achieving this behavior. Below, we explore the mechanistic basis for the generation of one such noncanonical histone modification pattern.

Noncanonical histone modification patterns represent coherent responses

Do nucleosomes that move to underrepresented regions of modification space reflect a biologically coherent response, or are these nucleosomes “aberrantly modified” based on accidental genomic juxtapositions between overlapping gene control programs? More specifically, do unusual modification patterns occur specifically in association with genes sharing a common regulatory strategy?

Searching a compendium of gene set annotations against the set of nucleosomes that explore non-canonical modification patterns revealed enrichment in multiple gene sets. For example, both ribosomal biogenesis genes (RiBi) and ribosomal protein genes (RPGs) are highly-expressed during mid-log growth and strongly repressed by diamide, and both were enriched with non-canonical modification patterns during stress, although they exhibit distinct trajectories (**Figures 4.5G-H** and **4.5L-M**). +1 nucleosomes of RiBi genes begin with high levels of K4me3 and K18ac, and in response to stress show rapid and dramatic loss of K18ac but much slower changes in H3K4me3, leading a large number of them to the K4me3 high/K18ac low state. In contrast, +1 nucleosomes of RPGs start in an extreme region of this 2D space as nucleosomes with the strongest signal for H3K18ac. They then show a transient increase in H3K4me3, followed by mild H3K18 deacetylation. This difference is consistent with the fact that repression of these two groups of genes involves different pathways (WEINER *et al.* 2012).

These and other examples (**Figures 4.5G-H** and **4.5M**) suggest that our approach identifies rare, but biologically meaningful, cases where regulatory features of specific groups of genes leads their chromatin transactions to differ from the standard pathways for gene induction/repression used by most genes.


A
Signal Signal
Time
X X
Y Y
transient
stable
B
Unexplained percentage
Percent variance explained
Principal components
0'
4'
8'
15'
30'
60'
C
0'
8'
30'
60'
PC2
PC1
D
H3K4me3
H3K18ac
Time
a
b
c
d
E
0'
30'
H3K4me3
H3K18ac
G
RiBi +1
RPG +1
F
H3K56ac
Htz1
H
RiBi +1
2-fold up +1

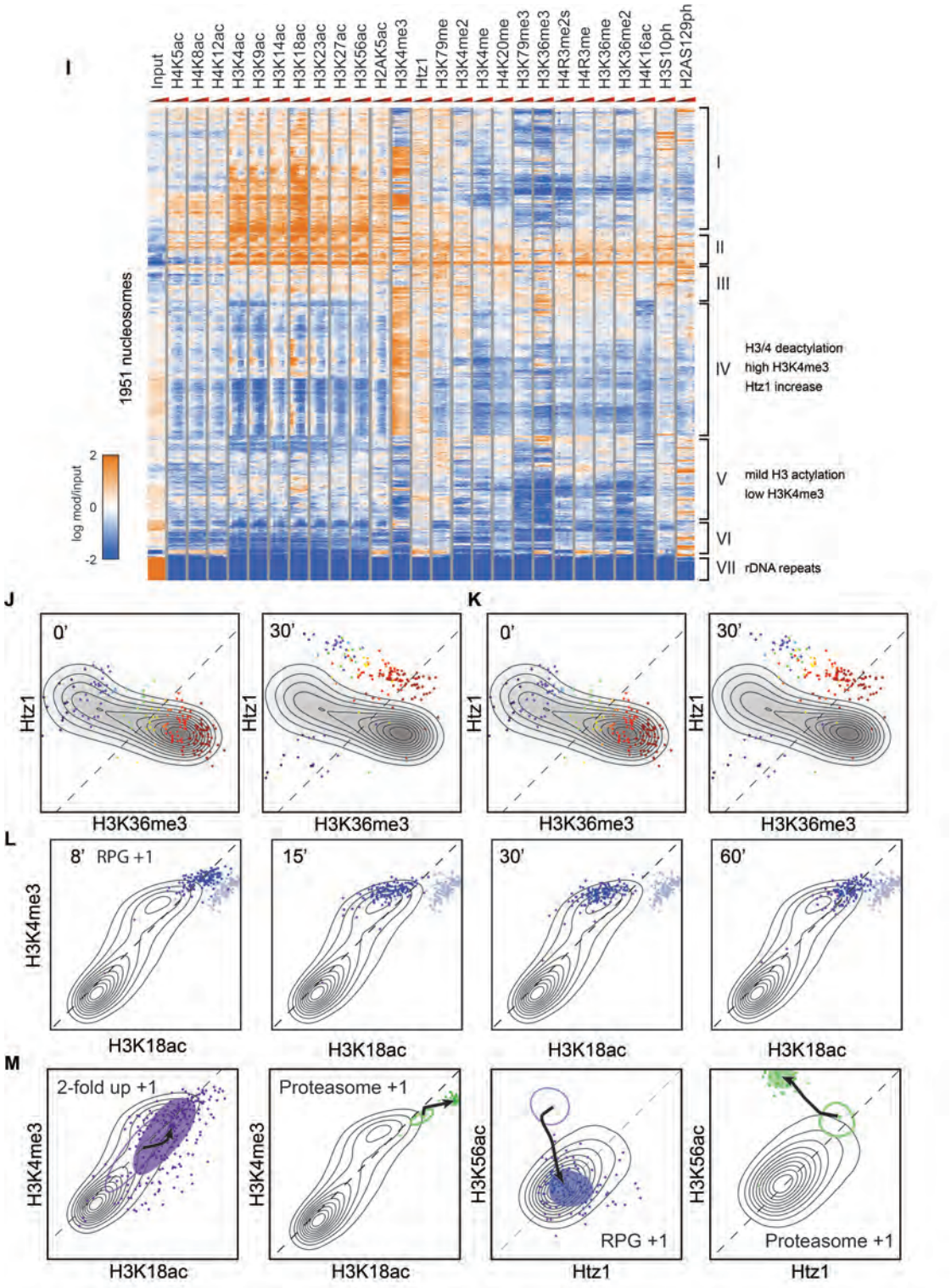


Figure 4.5. Changes in histone modification space during stress

(A) Schematic showing one potential mechanism leading to increased combinatorial complexity during a transient response. Briefly, if two histone marks are correlated but exhibit different relative response kinetics, then early during a change in transcription the nucleosomes will carry the rapid mark, but lack the lagging mark. **(B)** Principal Component Analysis of all 26 histone modifications. Percent variance explained for different time points. **(C)** Transient population of low-density modification space during stress. Density of nucleosomes across the first two Principal Components at the indicated diamide time points. Arrows show regions that are more highly populated from t=8 to 30 than during midlog growth. **(D)** Visualization of histone modification trajectories. Contour map shows the predominant locations of nucleosomes in the indicated 2D modification space at t=0. Arrows indicate the paths of two specific nucleosomes during the diamide time course. **(E)** Transient population of new regions of histone modification space. Left panel: 2D contour map for nucleosomes at t=0 for H3K4me3 and H3K18ac. Nucleosomes that will fall significantly (**Methods**) outside this contour during stress are color-coded according to their location at t=0. Right panel: the t=30 locations of nucleosomes that move to rare regions, with the t=0 contour. **(F)** As in (**E**), but for Htz1 and H3K56ac. **(G-H)** Coherent groups of nucleosomes account for the unusual nucleosomes during stress. Trajectories for specific sets of nucleosomes as indicated, with the t=0 domain marked by an empty oval, and the stress domain marked by points and a filled oval. **(I-M)** Nucleosomes traversing non-canonical combinations. **(I)** Heatmap showing input levels and modification levels (relative to input) for 1915 nucleosomes that leave the high-density region in the 26-dimensional space. Several prominent clusters are noted. **(J-K**). Movements of nucleosomes through 2D modification space. **(J)** The anticorrelation between H3K36me3 and Htz1 (and H2AK5ac) was violated by a number of nucleosomes at late timepoints. **(K)** Nucleosomes that transiently gain the repression-related H2AS129ph despite carrying high levels of H3K36me3 throughout the time course. **(L)** Movement of +1 nucleosomes of Ribosome Protein Genes in the H3K18ac/H3K4me3 space. Blue dots show nucleosomes at the relevant time point, gray dots show nucleosomes at t=0. **(M)** Trajectories for specific sets of nucleosomes are shown, with the t=0 domain being shown as an empty oval, and the stress domain shown as points and a filled oval (as in **Figures 4.5G and H**).

Dynamics of chromatin responses reveal subtle distinctions between histone marks

How do rare histone modification states become populated in response to stress signals? In the case of H3K4me3/H3K18ac, the transient uncoupling of H3K4me3 status and H3K18ac levels appears to result from a difference in the kinetics of each modification's response to transcriptional reprogramming – H3K18 deacetylation is rapid and occurs over genes subject to both short and longer-term repression, while H3K4 demethylation significantly lags deacetylation and is specific to longer-term gene repression.

To interrogate the dynamic behavior of individual histone modifications, we modeled the time course of histone modification changes at each nucleosome and extracted the time to half-maximal response ($t_{1/2}$), and the amplitude of maximal response (h) (**Figure 4.6A**). Our model interpolation enables continuous assignment of $t_{1/2}$ times across our time course (**Figure 4.6B**), allowing us to compare differences in kinetic behavior between similar modifications. We also estimated kinetic parameters for mRNA abundance changes (GASCH *et al.* 2000) – similar results are obtained with analyses based on Pol2 ChIP-chip data (KIM *et al.* 2010).

Comparison of average timings of different modifications revealed a range of behaviors (**Figure 4.7H**). For example, acetylation marks tend to change more rapidly than methylation, with gene body marks H3K36me3 and H3K79me3 changing later than the majority of other modifications. As the dynamics of

changes can depend on the location of the nucleosome on a gene and the transcriptional response of the gene, we focused on kinetics of those modifications whose change is correlated/anti-correlated with transcription (**Figure 4.4H**), and analyzed modification dynamics associated with the relevant gene region (**Figure 4.6C**). Interestingly, modification dynamics differ significantly between induced and repressed genes. For example, loss of H3K4me3 at the 5' end of repressed genes is more rapid than establishment of the same mark over induced genes, presumably reflecting the time required for successive addition of up to three methyl marks. In contrast, establishment of H3 tail acetylations (with the exception of H3K27ac) at promoters of induced genes is more rapid than corresponding deacetylation at repressed genes. Modifications that are anti-correlated with transcription tend to occur later than acetylation marks, with more widely distributed $t_{1/2}$ values. Among these, we notice a significant difference between the timing of H2AS129ph increase at repressed genes and its decrease at induced genes.

These results show clear differences in the timing of events based on their location and function. Changes at the 5' end of genes, which are associated with either promoting or inhibiting initiation, tend to occur early in the response, while changes at the gene body, which are associated with elongating transcription, tend to appear later.

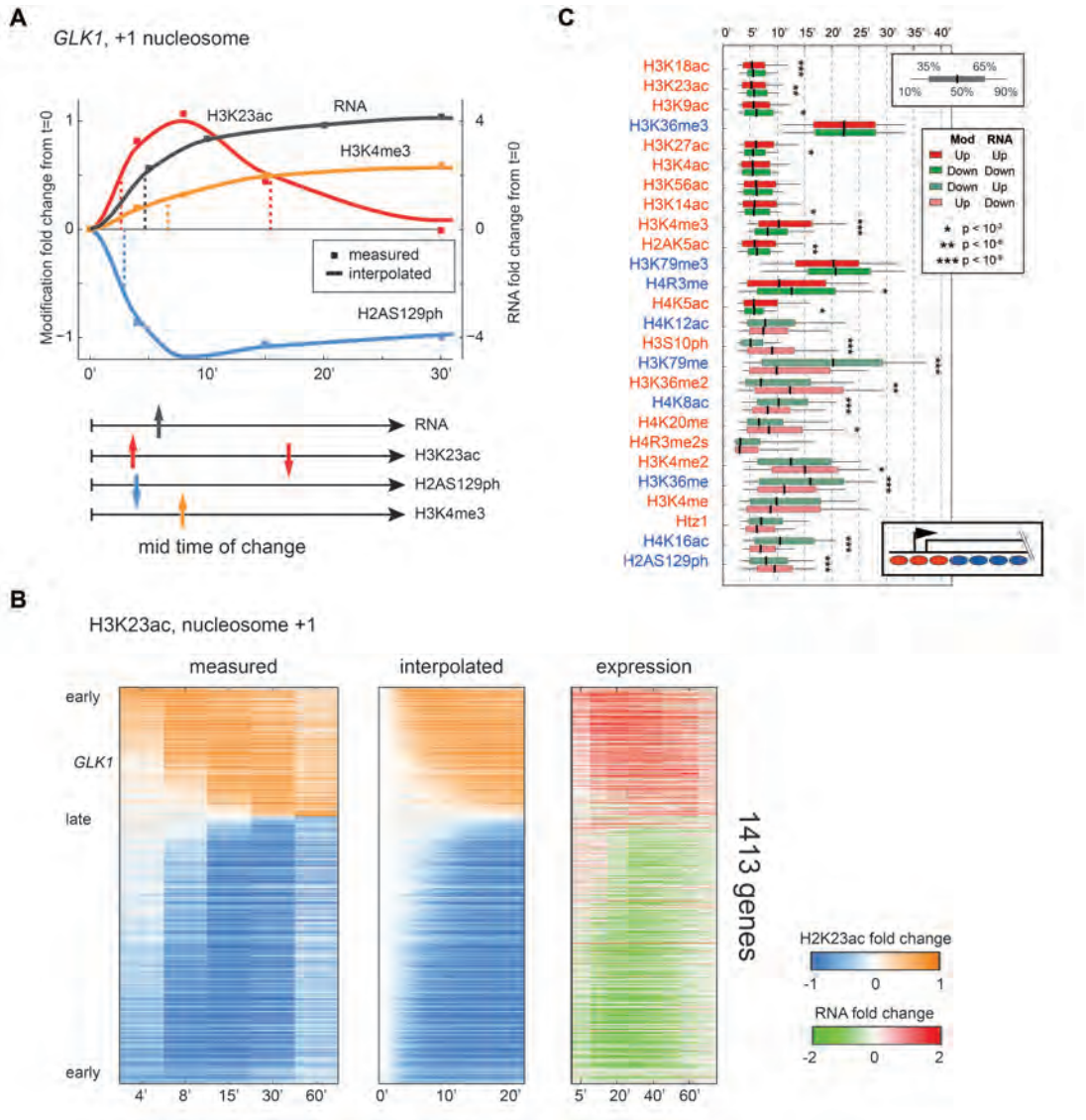


Figure 4.6. Analysis of histone modification dynamics

(A) Extraction of kinetic parameters from time course data. RNA abundance and the indicated modification levels for the *GLK1* +1 nucleosome. For each time course we extracted the maximal response (h) and the time to half-maximal response ($t_{1/2}$). **(B)** Comparison of measurements with extracted kinetic data, with rows showing individual genes. Left panel: time course data for H3K23ac levels at the +1 nucleosome sorted by $t_{1/2}$; Middle panel: interpolated data; Right panel: mRNA abundance changes. **(C)** Genome-wide kinetic offsets for up- and down-regulated genes. For each modification, boxplot of the $t_{1/2}$ is shown for up- or down-regulated genes, as indicated.

Modification cascades in transcriptional reprogramming

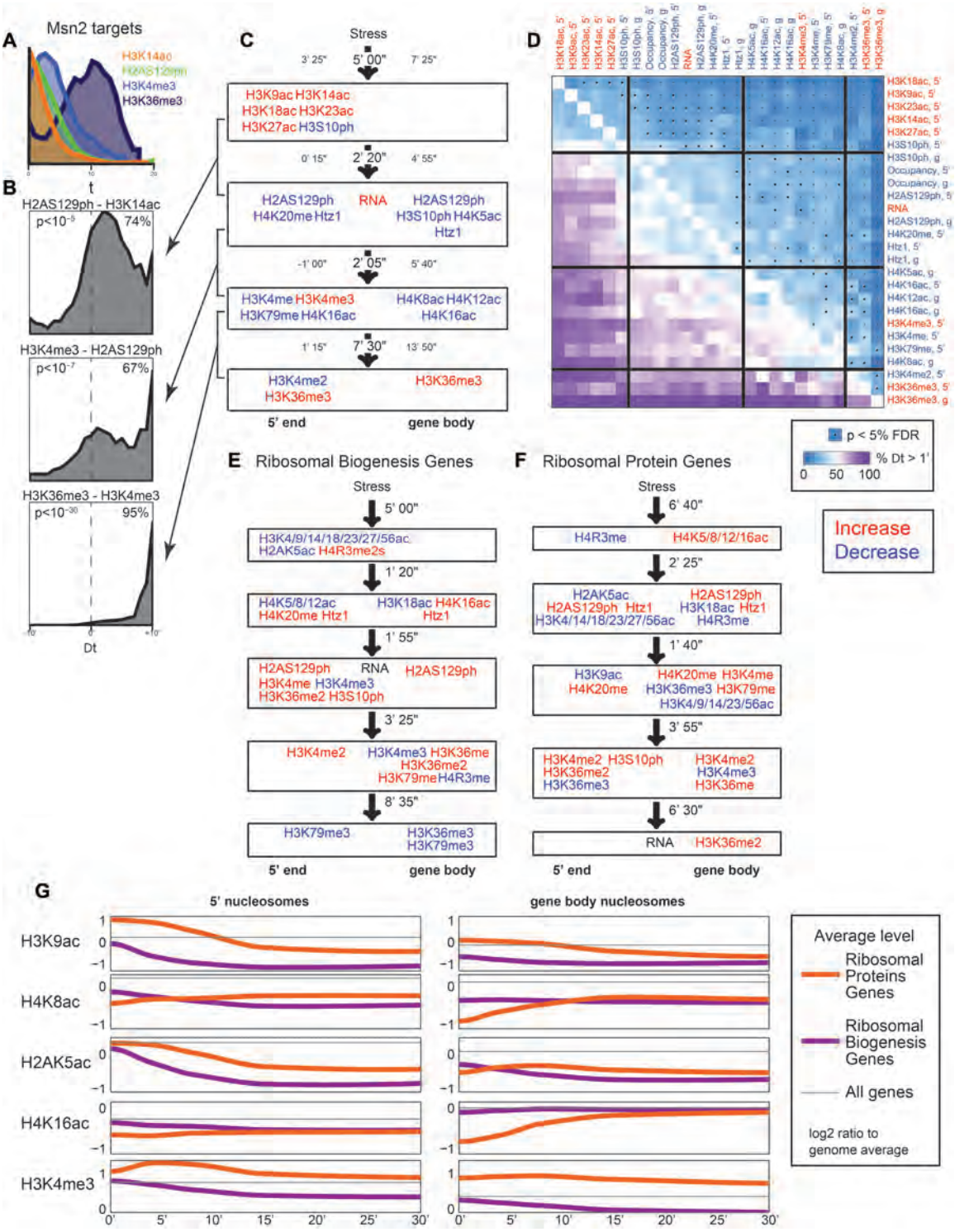
Finally, we turn to single gene analysis of dynamics to gain a more biologically relevant picture of chromatin events in transcriptional reprogramming. To systematically analyze the ordering of 26 histone marks and transcription events, we calculated timing differences (Δt**)** between all pairs of marks for every gene in the yeast genome, revealing behaviors not apparent in individual plots of timing distributions. For example, analysis of the 5' nucleosomes of Msn2-dependent induced genes revealed a subtle difference in the timing of H3K14ac and H2AS129ph across all genes as a group (**Figure 4.7A**). However, examining the $t_{1/2}$ difference between these marks on a gene-by-gene basis revealed a striking and significant timing difference (**Figure 4.7B**). In this example, H3K14 acetylation precedes a change in H2AS129ph at the majority of Msn2 target genes (74%). These timing differences may suggest ordered recruitment of chromatin regulators in response to stress.

Gene-by-gene analysis of repressed and induced genes (**Figures 4.7I-J**) recovers cascades of events occurring during changes in transcription, which mainly recapitulates the order of events we observed above (**Figure 4.6C**). The dominant pattern in both analyses reflects 5' acetylation marks changing prior to changes in mRNA abundance, with gene body methylation following. By assessing significant kinetic differences between pairs of events (**Methods**), we

can identify at least four clearly distinct temporal stages in the chromatin response to transcriptional activation (**Figures 4.7C-D**).

Beyond aggregating our gene-by-gene kinetic offsets into gross genesets consisting of all up- or down-regulated genes, these data can identify gene sets that have significantly coherent temporal event cascades. For example, the analysis above (**Figures 4.5E-G)** suggested that histone modifications might exhibit distinct kinetic behaviors at the Ribosomal Biogenesis genes and Ribosomal Protein genes. Indeed, although both groups are repressed, they show rather different histone modification cascades (**Figures 4.7E-G** and **4.7K**). Several substantial differences can be appreciated between these highly repressed genesets, including modifications that change over different timescales (e.g., H3K9ac and other H3 acetylation marks), and modifications exhibiting different magnitudes of change (e.g., H3S10ph and H2AK5ac). More interestingly, several modifications exhibit opposite behaviors: for example, H4 N-terminal lysines are deacetylated at RiBi, but not RPG, 5' ends, and conversely are strongly acetylated only over RPG gene bodies. Finally, we confirm our previous finding (WEINER *et al.* 2012) that H3K4me3 is transiently induced at the 5' ends of RPGs prior to being lost later during RPG repression. Curiously, most of these differences in modification profiles during stress reflect initial differences between RiBi and RPGs at t=0, as for most modifications the two groups are more similar in the "off" state (at t=30) than in the "on" state (t=0) (**Figure 4.7G**).

This suggests that some of the observed differences are due to different mechanisms involved in their mid-log transcription.



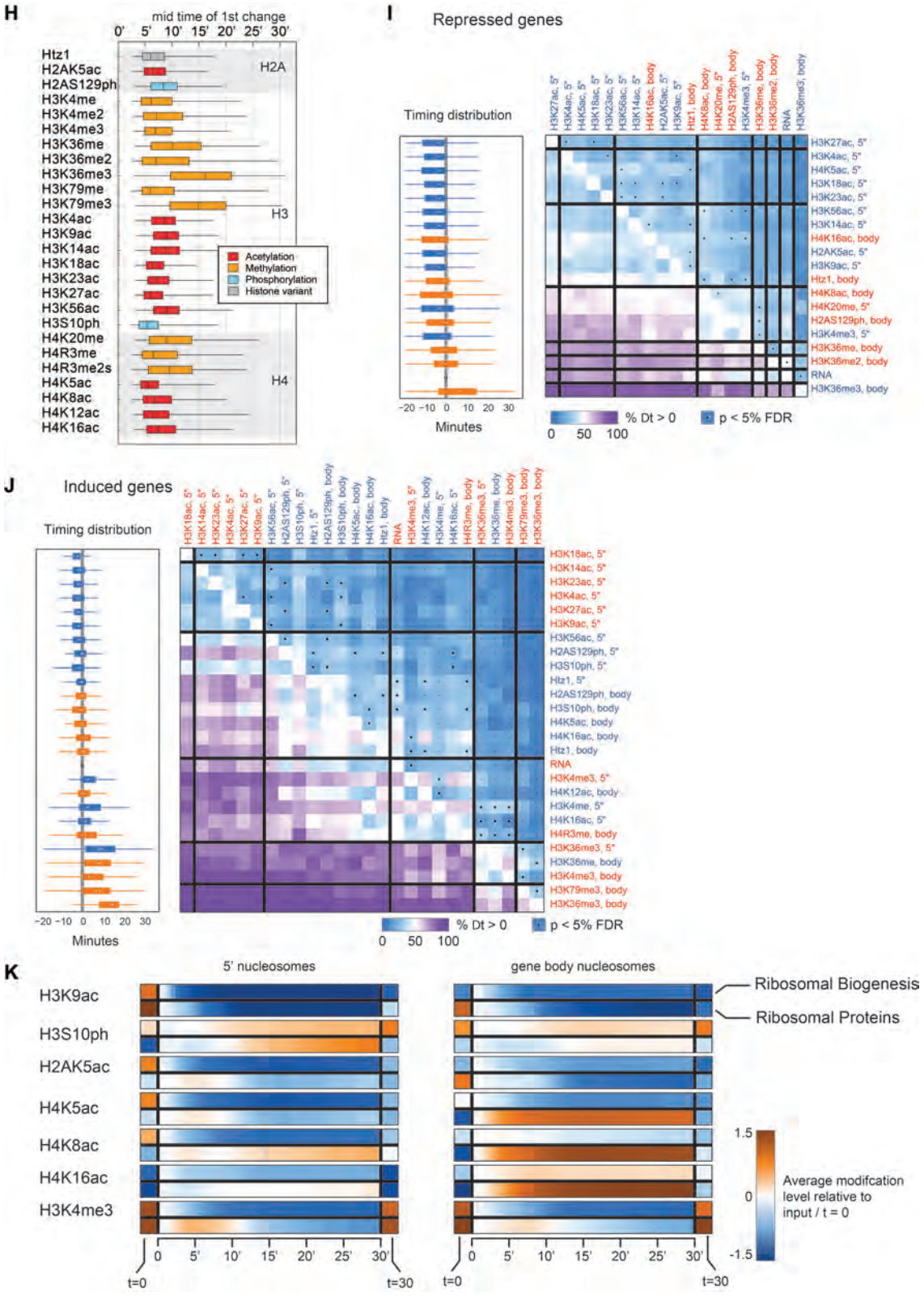

H
mid time of 1st change
0' 5' 10' 15' 20' 25' 30'
Htz1
H2AK5ac
H2AS129ph
H3K4me
H3K4me2
H3K4me3
H3K36me
H3K36me2
H3K36me3
H3K79me
H3K79me3
H3K4ac
H3K9ac
H3K14ac
H3K18ac
H3K23ac
H3K27ac
H3K56ac
H3S10ph
H4K20me
H4R3me
H4R3me2s
H4K5ac
H4K8ac
H4K12ac
H4K16ac
H2A
H3
H4
Acetylation
Methylation
Phosphorylation
Histone variant
I
Repressed genes
Timing distribution
-20 -10 0 10 20 30
Minutes
% Dt > 0
p < 5% FDR
0 50 100
J
Induced genes
Timing distribution
-20 -10 0 10 20 30
Minutes
% Dt > 0
p < 5% FDR
0 50 100
K
5' nucleosomes
gene body nucleosomes
H3K9ac
H3S10ph
H2AK5ac
H4K5ac
H4K8ac
H4K16ac
H3K4me3
Ribosomal Biogenesis
Ribosomal Proteins
0 5' 10' 15' 20' 25' 30'
t=0
t=30
1.5
0
-1.5
Average modifcation level relative to input / t = 0

Figure 4.7. Cascades of chromatin events differ between genesets

(A) Distribution of $t_{1/2}$ values for the four indicated marks for all MSN2-induced genes. **(B)** Gene-by-gene analysis for differences in modification onset times. The distribution of the difference in $t_{1/2}$ is calculated for all individual genes in the MSN2-dependent geneset for the indicated modification pair. **(C)** Four "epochs" in the MSN2 induction cascade. Groups of histone modification changes: modifications in each group roughly co-occur, but differ significantly in timing in pairwise comparisons from the other groups. For each box, the mean and 25th and 75th percentile values are shown for the distribution of differences in $t_{1/2}$ between modifications in adjacent boxes. **(D)** Heatmap showing all pairwise comparisons for MSN2-dependent upregulated genes. Each row/column represents a modification and a genic location (**5'** end, or **g**ene body) that changes coherently for MSN2-upregulated genes. Heavy lines show demarcation for the boxes summarized in (**C**). **(E-F)** Summary diagrams, as in (**C**), for RiBi genes and RPGs, as indicated. **(G)** Interpolated time course data for RiBi genes and RPGs for 30 minutes of stress response. The shown modification levels are average over genes in each group of the log2 ratio to genome-wide mean at t=0. **(H-K)** Timing of changes. **(H)** t1/2 global by mod. Moreover, there are noticeable differences between specific acetylation marks, as H3K18ac, H3K23ac, H3K27ac, and H4K5ac change earlier in the response, while H3K56ac and H4K16ac are slower. Changes in H3S10ph and Htz1 also change as rapidly as the earliest acetylations, while H2AS129ph is relatively slow. **(I)** Repressed genes and **(J)** Induced genes. Right: matrix of relative timing as in **Figure 4.7D.** Left: box-plot of $t_{1/2}$ for each modification relative to RNA $t_{1/2}$. **(K)** Interpolated time course data for RiBi genes and RPGs for 30 minutes of stress response (as in **Figure 4.7G**), shown in heat map representation. The left-most and right-most cells denote level relative to genome wide mean at t=0, 30, respectively. The middle row shows changes relative to t=0.

DISCUSSION

This study represents the deepest characterization to date of the primary structure of the yeast chromatin landscape, with nucleosome positioning and 26

histone modifications mapped at nucleosome resolution genome-wide under standard conditions (growth in YPD), and during five time points of a well-characterized stress response. The data reproduce essentially all known characteristics of yeast histone modification localization, and provide further insights into histone modification biology.

Steady-state patterns of histone modifications

Analysis of histone modifications in actively growing “midlog” yeast confirms and extends a great deal of prior knowledge. In general, histone modification patterns exhibit little combinatorial complexity, as repeatedly observed in many organisms (RANDO 2012). Essentially, we identify three major features that explain nucleosomal modification patterns. First, chromosome context impacts histone phosphorylation states – H3S10ph marks broad pericentric domains, while H2AS129ph marks subtelomeric domains. Second, the process of transcription leaves a massive footprint on chromatin, with enzymes carried by the initiation or the elongation form of RNA Pol2 being responsible for the majority of the variation in histone modifications across the genome. Finally, replication-independent histone replacement – which is modestly correlated with, and affected by, transcription – is responsible for deposition of histones carrying marks such as H3K56ac and lacking H3K79me3 and other marks. The roles of these factors in chromatin structure are all conserved to varying extents in other

organisms, with many other organisms exhibiting additional elaborations such as the H3K9 and H3K27 methylation-dependent repressive mechanisms.

Rules of chromatin marks are broadly maintained during reprogramming

Our analysis of a stress response reveals that the relationships observed between modifications and transcription levels in mid-log growth are generally maintained during transcriptional reprograming – modifications that are correlated with transcription rate in steady-state conditions also increase during gene activation and decrease during gene repression. These observations imply that, broadly, the mechanisms that maintain the chromatin modification landscape in mid-log growth are the same or similar to the ones involved in changes during stress-induced transcriptional reprogramming. Thus, we argue that while the shift in cellular context from mid-log growth to stress response changes the transcription program (eg from TFIID-dominated to SAGA-dominated gene regulation), it does not change the rules governing the deposition and maintenance of chromatin marks. This suggests that the mechanisms that deposit most transcription-related marks are generic to transcriptional machinery rather than to the context in which it is activated or repressed.

Combinatorial complexity during transcriptional reprogramming

The extensive crosstalk between transcription and histone modifications results in limited histone modification complexity. Despite the potential for widespread network motifs such as incoherent feedforward loops to generate transient combinatorial complexity in the histone modification network, we only observed a modest increase in combinatorial complexity during the peak of the stress response. Overall, we found that 3% of nucleosomes move into normally sparse regions of the histone modification space in response to diamide stress, despite ~60% of all nucleosomes moving within this space during the stress response (**Figure 4.5**).

A number of mechanistically distinct processes could transiently violate steady-state histone modification correlations, including complex crosstalk loops, kinetic offsets between correlated marks, or population heterogeneity in gene induction. In the case of H3K4me3/H3K18ac, these traces reveal two major behaviors of nucleosomes that are rapidly deacetylated at H3K18 while still H3K4-methylated: a subset recover to the original modification status as yeast adapt to stress and reactivate transiently-repressed genes, while another group of these nucleosomes instead lose H3K4me3 due to ongoing repression of the associated gene. In other words, transient uncoupling of H3K4me3 status and H3K18ac levels results from a difference in the kinetics of each modification's response to transcriptional reprogramming – H3K18 deacetylation is rapid and occurs over genes subject to both short and longer-term repression, while H3K4

demethylation significantly lags deacetylation and is specific to longer-term gene repression.

Ordered waves of histone modifications during transcriptional reprogramming

Although transcription-related modifications increase and decrease in expected ways upon changes in transcription, we see marked differences in the timing of these changes. In general, acetylation changes at the 5' of genes appear early in the transcriptional response, while gene body methylation occurs more slowly. Similar timing differences were recently observed during yeast exit from starvation state (MEWS *et al.* 2014). Here, the timing of changes depends on the transcriptional program, as different co-regulated gene sets exhibit distinct cascades of modification changes. Understanding whether these cascades reflect independent events with different temporal delays, or linear chains of dependent events, will require further experiments with denser temporal samples and genetic or drug interventions. Our analysis provides an inventory of the relevant time scales and the representative modifications to follow in such detailed experiments.

Furthermore, although in general modification changes are generic, there are subtle differences in the timing and intensity of changes during repression/induction of differently regulated genesets. This observation suggests that regulatory mechanisms alter the footprints made on the chromatin

modification landscape. Most notably, ribosomal protein genes and ribosomal biogenesis genes, both of which are strongly growth-related in expression, exhibit significant differences in chromatin dynamics during repression.

Toward a comprehensive view of chromatin dynamics

The dataset and analysis presented here provide a detailed and comprehensive view of chromatin state in yeast and how it responds to a massive transcriptional reprogramming event. Chromatin changes are intimately connected to transcriptional changes, occurring with clearly defined ordering relative to transcription. Although such observational data does not provide evidence of causality, it provides a rich resource for evaluating potential pathways and suggesting interventional experiments to further resolve the myriad interactions between chromatin marks and transcription.

Materials and Methods

Cell culture and MNase-ChIP

Wild-type yeast (BY4741) cells were grown in six flasks of 400 mL YPD to mid-log phase (OD600=0.55) shaking (220 rpm) at 30°C. Cells were treated with diamide (1.5 mM) and fixed at 0, 4, 8, 15, 30, and 60 min with 1% formaldehyde for 15 min. Cell pellets were harvested, washed by water, and subjected to bead-beating, MNase digestion, and chromatin immunoprecipitation (Liu *et al.* 2005).

Library preparation and sequencing

Multiplexed libraries were prepared using HT-ChIP (BLECHER-GONEN *et al.* 2013). Libraries were sequenced on an Illumina HiSeq-1500 (50bp single end sequencing) to obtain ~5-10 million aligned reads per sample. Reads were mapped to the *S. cerevisiae* sacCer3 assembly using 'bowtie2' with default parameters, and only tags that uniquely mapped were used for further analysis.

Data processing and normalization

Except for metagene views, all analyses where performed on nucleosome discretized and occupancy-normalized data. Using these values, we estimated the log ratio of ChIP coverage compared to input in each sample. Values within each time series (antibody X time points) were quantile-normalized using MATLAB (version R2013a) quantilenorm function.

Regression and sparse regression

We used multiple linear regression analysis to reconstruct histone modifications levels from a collection of features: Nucleosome position, mid-log occupancy (input), NET-seq (CHURCHMAN AND WEISSMAN 2011) coverage both in sense and anti-sense directions, turnover rate (DION *et al.* 2007), replication time (RAGHURAMAN *et al.* 2001), and log of distance to nearest telomere or centromere. Since position is a discrete feature, we estimated a different regression model for each nucleosome position.

To use histone modifications to predict genomic features we applied sparse linear regression (MATLAB's lasso procedure) with 5-fold cross validation to estimate mean squared error (MSE). We fitted the Lambda parameter value (nonnegative regularization parameter) with the minimum MSE using the default Lambda scan method.

Detecting nucleosomes at low density regions

To investigate the 26-dimensional modification space, we used kernel density estimation, with bandwidth determined by cross-validation. To mark nucleosomes that arrive at low-density regions, we take the 0.1% quantile of the density at mid-log as our threshold for the definition of "low" density at all other time points. We used the same approach to detect low-density regions in pairwise dimensional projections of the data.

Fit and $t_{1/2}$ estimation

We use a non-parametric approach using multiple leave-one-out estimates to interpolate modification changes at each nucleosome and evaluate the accuracy of the interpolation. Given the estimate, the peak change, h, is defined as the point in time which has the maximal absolute change, relative to t=0. We define $t_{1/2}$, as the time at which the estimated response reaches half the peak change. For each nucleosome and each modification, we use a permutation test to evaluate whether the observed time trajectory is non-random**.**

Timing of events in a gene set

To identify coherent events at the level of genes and gene sets we partition nucleosomes in each gene to 5' (-2,-1,+1,+2) and gene body (>= +3). For each gene, we average only the timing statistics of *coherent* nucleosomal changes (as defined above, 25% FDR) across these nucleosome sets to obtain events per gene and position (5'/gene body) for each modification, and for occupancy, and RNA levels. Each of these events has its peak change and $t_{1/2}$. Next, for given set of genes, each of the above events is considered coherent in the gene set, if: 1) at least 40% of genes show a coherent response. 2) 75% of those responses changed in the same direction ("up"/"down"), and 3) the distribution of changes was significantly non 0-centered (t-test with 5% FDR).

Ordering events

We define the precedence of one event, A, over the other, B, with respect to a gene-set G as the fraction of genes in which event A is preceded by event B by at least one minute. For each such coherent event pair, we performed one-sided t-tests (with 5% FDR). Pairs that pass the test are defined as significant precedence relations. The set of these pairs define the Timing of Events (TOE) graph w.r.t to the gene-set G. We cluster events to "comparable" clusters**.**

CHAPTER V

Discussion

Chromosome conformation at single-nucleosome resolution

The compaction and organization of the physical genome have wide-ranging causes and consequences for genomic functions. In eukaryotes, the characteristic “beads-on-a-string” structures are depicted as the first level of genome compaction, with the “beads” being separated by relatively accessible linker DNA at an average ~147 bp of nucleosome repeat length (FELSENFELD AND GROUDINE 2003). This primary structure of chromatin has been well-characterized, with multiple solved crystal structures for the nucleosome (LUGER *et al.* 1997; KORNBERG AND LORCH 1999), and a plethora of genome-wide studies (HUGHES AND RANDO 2014) that identify the locations and the chemical modifications of individual nucleosomes across the genome in various organisms, sometimes at single nucleotide-resolution (BROGAARD *et al.* 2012). In general, well-positioned nucleosomes are located at the borders of nucleosome depletion regions (NDRs), with chemical marks enriched for multiple acetylations on histone tails and di/trimethylation on H3K4. Precise nucleosome positioning gradually decays with distance, as fuzzy nucleosomes are widely observed at 1 – 2 kb downstream of the transcription start sites (HUGHES AND RANDO 2014). However, the next level of chromatin folding is increasingly subjected to debate. Do regular chromatin structures (e.g., the 30 nm chromatin fiber) exist *in vivo*?

What kind of helical paths (e.g., zigzag or solenoid) are more entropically favorable? How do these chromatin structures manage to fold into higher-order structures, and how do these structures relate to genomic functions?

The Chromosome Conformation Capture (3C) family of techniques have revolutionized our understanding of chromosome conformation beyond the 30 nm chromatin level (DEKKER *et al.* 2002; SIMONIS *et al.* 2006; ZHAO *et al.* 2006; DOSTIE AND DEKKER 2007; LIEBERMAN-AIDEN *et al.* 2009). The methods measure interaction probability between genomic loci that are crosslinked together *in vivo*, and the resulting measurements of ligated products can be translated to spatial distance and physical models of chromosome conformation in the nucleus. The 3C-related methods with increasingly fine resolution and improved experimental/analytical workflows have redefined our understanding of the organizational features of eukaryotic genomes, zooming in on the order of chromosome folding including full scale chromosomal territories, multiple-megabase compartments (LIEBERMAN-AIDEN *et al.* 2009), hundred-kilobase topologically-associating domains (TADs) (DIXON *et al.* 2012; NORA *et al.* 2012), and various sizes of chromatin loops (PHILLIPS-CREMINS *et al.* 2013; SOFUEVA *et al.* 2013; RAO *et al.* 2014). These chromatin structures play key roles in a large number of nuclear processes. For example, gene loops are proposed to enforce transcriptional activity and directionality in yeast (O'SULLIVAN *et al.* 2004; ANSARI AND HAMPSEY 2005; TAN-WONG *et al.* 2009), TADs correlate with many tightly regulated operations in mammals such as replication timing (TAKEBAYASHI *et al.*

2012; POPE *et al.* 2014) and developmental clusters (SYMMONS *et al.* 2014; CHANDRA *et al.* 2015; LUPIANEZ *et al.* 2015; FRANKE *et al.* 2016), and LADs enclose silencing genes at the nuclear periphery during development (GUELEN *et al.* 2008; REDDY *et al.* 2008; KIND *et al.* 2013; KIND *et al.* 2015). Many biological consequences and diseases have been identified in relation to chromatin domains, such as cancer (GROSCHEL *et al.* 2014; LE DILY *et al.* 2014; ACHINGER-KAWECKA AND CLARK 2017), aging (CHANDRA *et al.* 2015; CRISCIONE *et al.* 2016a; CRISCIONE *et al.* 2016b), and autoimmune diseases (MARTIN *et al.* 2016; MCGOVERN *et al.* 2016). However, the suboptimal resolution of the current methods, typically yielding few kilobase average fragment size, leaves a "blind spot" on our full understanding of chromatin organization, including the length scale relevant to secondary chromatin folding such as 30 nm fibers, gene loops, or nucleosome stacking. Filling in the missing pieces of chromatin folding at the scale of individual genes or ~1 – 10 nucleosomes will also uncover how chromatin associates with local transcriptional regulation, nucleosome positioning, and histone modifications in unprecedented detail.

Concepts of Micro-C protocol

In Chapter II, we developed an innovative 3C-based method – Micro-C – in which chromatin is fragmented into mononucleosomes using micrococcal nuclease (MNase), thus essentially enabling nucleosome-resolution maps of chromosome folding. Each step in the protocol uniquely contributes to generating high-quality

of Micro-C data. First, precise control of chromatin fragmentation is the most critical step in the protocol. Underdigestion results in overrepresentation of adjacent nucleosomes which contaminate the pool of ligated products (particular in the “IN-IN” dataset), while overdigestion leaves nucleosomal ends too short to be ligatable (data not shown). Ideal digestion results in an optimal range of 90 – 95% monomers and 5 – 10% dimers. Second, to prepare ligatable ends, we used the 3’ to 5’ exonuclease activity of T4 DNA polymerase or DNA polymerase Klenow in the absence of dNTP to chew-in one strand of DNA, which stops once the enzyme hits the border of a crosslinked nucleosome. The 5’ to 3’ polymerase activity of these enzymes instantly overrides their exonuclease function upon providing dNTPs, which then generates ligatable blunt ends (**Figure 2.1A**). Third, the two-step size-selection process, with the first selection after ligation and the second selection after library construction, ensures only chimeric DNA from genuine nucleosomal contact will be sequenced (**Figure 2.1A**). We then generated high-coverage chromatin interaction maps at single-nucleosome resolution for the budding yeast *S. cerevisiae*, along with the proof-of-principle controls, including no-crosslinking, no-ligation, and mix-n-ligation with *K. lactis*, proving the extraordinary specificity and sensitivity of Micro-C protocol (**Figures 2.1B-D**). It worth noting that the materials of the no-crosslinking control were prepared from reverse crosslinked DNA because theoretically this yields the maximal random contact frequency over native chromatin (BELTON *et al.* 2012; LAJOIE *et al.* 2015). A recent Hi-C study on native chromatin argued that the

intrinsic forces between proteins and proteins or proteins and DNA/RNA are sufficient to maintain some extent of stable interactions rather than yielding completely random contacts (BRANT *et al.* 2016).

Micro-C efficiently captures short-range chromatin interactions, but provides very weak signals for well-validated long-range interactions (e.g., telomere-telomere and centromere-centromere interactions) (**Figures 2.2E-F**). Curiously, this seems to be a common effect of increasing resolution in 3C methods: Hi-C studies using 4-cutters are less effective at identifying long-range interactions than are 6-cutter-based Hi-C methods. The reason for this is unclear; however, as the vast majority of Micro-C ligation products occur within ~1 – 2 kb, it seems likely that being unable to sequence these products in traditional 4 kb resolution assays allows observation of very rare long-distance ligation products. It is possible that fine-tuning the Micro-C protocol (e.g., crosslinking or ligation) may improve the capture of long-range interactions (**Chapter III**).

Chromatin domains and boundaries

Analysis of Micro-C results revealed abundant self-associating domains typically spanning 1 – 5 genes (~2 – 10 kb) (**Figures 2.2A-F**), which resemble chromatin domains (TADs) found in mammals but previously were unidentifiable in organisms with small-genome. Prior Hi-C data indicated that chromatin domains are physically demarcated by the boundaries/insulators, whose activity tightly associates with active genes and epigenomic marks, architectural proteins

(CTCF, Cohesin, and Mediators), and repetitive elements (SINEs) in mammals as well to some extent in drosophila (DIXON *et al.* 2012; SEXTON *et al.* 2012). The increased resolution of Micro-C data allows us to further define the boundaries in yeast preferentially localized at promoters of highly-transcribed genes, enriched with the active epigenomic marks (e.g. acetylation on histone tails and H3K4me3), and regions of rapid histone turnover (**Figures 2.3A-C**). Binding of the complexes for RSC ATP-dependent chromatin remodeling and Cohesin loading at the boundary regions (**Figures 2.3D-E**) strongly suggests that they help facilitate the boundary forming process, as the working model posits that the RSC complex first binds onto the stiff poly A/T tracks in promoter regions, evicts local nucleosomes, and then recruits Cohesin-loading complex. RSC and Cohesin loader have synergetic effects on transcription regulations (LOPEZ-SERRA *et al.* 2014), which further argues that the formation of boundaries is tightly associated with transcriptional regulation. Although Cohesin is recruited by Cohesin loaders, their biological functions may differ from each other: FISH imaging and ChIP assays show differing DNA binding signatures (CIOSK *et al.* 2000), and Cohesin-depletion and loader-depletion mutants show inconsistent effects on chromosome conformation (SEITAN *et al.* 2013; SOFUEVA *et al.* 2013; ZUIN *et al.* 2014; SCHWARZER *et al.* 2016). Moreover, evaluation of a range of MNase digestions of chromatin has negligible effects on global interactions and boundary callings. Underdigestion (with a visible trinucleosome band) or overdigestion (with a faint dinucleosome band) of chromatin prior to Micro-C

looks nearly identical to standard Micro-C data, arguing against the idea that boundary calling is biased by MNase digestion (**Figures 2.2L-M**).

Interestingly, self-associating domains tend to be conserved between species as scaled by gene numbers, rather than linear sequence. The genomic length of mammalian TADs (> 100 kb) (DIXON *et al.* 2012; RAO *et al.* 2014) or drosophila TADs (> 50 bp) (HOU *et al.* 2012; SEXTON *et al.* 2012) is much longer than in yeast, while the domains in these species cover roughly the same number of genes per domain. This disparity in length scale suggests that the primary determinant of chromosome folding may be the boundaries that separate compacted domains, rather than the internal structure of the domains themselves, as the detailed folding of the “beads on a string” within a given domain likely varies quite a bit between different organisms. In other words, we speculate that establishment of boundaries delimiting chromatin domains is the major driver of chromosome folding behavior, with the folding of the chromosome within each boundary-delimited domain not necessarily conforming to any regular secondary structure.

Gene loops vs. Gene crumples

Gene looping structures suggest direct contacts between 5’ and 3’ ends of gene, mediated by Pol2, general transcription factors (GTFs) (O'SULLIVAN *et al.* 2004; SINGH AND HAMPSEY 2007; MEDLER *et al.* 2011; MEDLER AND ANSARI 2015), activators (e.g., Mediators) (EL KADERI *et al.* 2009; MUKUNDAN AND ANSARI 2011;

MUKUNDAN AND ANSARI 2013), and terminator factors (ANSARI AND HAMPSEY 2005; MEDLER *et al.* 2011; AL HUSINI *et al.* 2013). They commonly have been proposed to facilitate transcriptional regulation such as gene activation or repression (YADON *et al.* 2013; LIU *et al.* 2016), recycling of transcription machinery (TAN-WONG *et al.* 2009), and promoter directionality (TAN-WONG *et al.* 2012). However, there is no evidence in the Micro-C data indicating that gene loops exist, at least not a ubiquitous structure *in vivo* (**Figures 2.1B** and **2.4A-B**). If they were to exist, the gene looping structure would appear as a spike signal at the corner of the gene box. However, we were only able to find heterogeneous interactions across each individual gene box, either by visual examination or in a metagene pile-up analysis, suggesting the presence of gene crumples or gene globules rather than gene loops. Several possibilities could result in the discrepancy between Micro-C data and prior restriction enzyme-based studies. First, although nucleosomes are capable of inhibiting restriction enzyme digestion of DNA, and, in principle, should prevent complete digestion of the genome, and bias cut sites to long linkers (such as those found at regulatory DNA), the aggressive detergent conditions typically used prior to restriction digestion in most Hi-C protocols results in complete or near-complete digestion of the genome in most studies (BELTON *et al.* 2012). The detergent conditions used for typical Hi-C methods presumably loosen histone-DNA contacts and potentially lead to dissociation of uncrosslinked proteins from DNA. Loss of nucleosomes during restriction digestion of chromatin in Hi-C could plausibly bias restriction enzyme-based

methods away from nucleosome-nucleosome interactions. From an experimental perspective, when chromatin is solubilized by detergent conditions prior to MNase digestion, as in the typical Hi-C protocol, the DNA ladder pattern is diminished, and the DNA loosened from nucleosomes is over chewed-in by MNase, as shown in the resultant size of DNA below 100 bp, even with a low level of MNase digestion. Second, previous studies suffered from suboptimal resolution, and an inability to conclude gene structures, as restriction enzymes often have only 2 – 3 cutting sites within a single gene in yeast regardless of gene length, thus leading to biases in interpretation of results with a limited number of ligatable ends (O'SULLIVAN *et al.* 2004; TAN-WONG *et al.* 2012). Third, the 3C protocol used in yeast gene loop studies includes a step to separate soluble and insoluble chromatin (ANSARI AND HAMPSEY 2005; TAN-WONG *et al.* 2012). Therefore, nucleosomes associated with relatively insoluble transcriptional machinery, such as the nucleosomes at the 5' and 3' ends of gene could be enriched preferentially in gene loop assays. Finally, the gene loop signals also could reflect the TF-TF interactions or transcriptional hubs. Another technical distinction between restriction enzymes and MNase is that restriction enzyme-based methods have the potential to identify direct TF-TF crosslinks, while the use of MNase digestion and gel isolation of dinucleosome-sized ligation products in Micro-C means that our method will only identify those interactions between transcription factors that also cause nearby nucleosomes to come into contact in a geometry amenable to DNA ligation (which presumably is common, but may

not be universal). However, efforts to specifically purify TF-TF interactions, by gel-purifying post-ligation products of <100 bp or 100-200 bp (average size of TF binding site is ~35 – 75 bp), resulted in massive overrepresentation of IN-IN DNA resulting from undigested DNA, and so were not pursued further. In sum, these are possible explanations for the differences in the structures of compacted genes between prior 3C studies in yeast, which reported “gene loop” structures, and the “gene crumple” structures observed within CIDs in this study.

Mutant effects on gene crumples

Levels of gene compaction are moderately anticorrelated to the transcription rate, as confirmed by measuring the changes in gene compaction under conditions of transcription inhibition and diamide oxidative stress (**Figures 2.2E-H**). Gene compaction increases for transcriptionally downregulated genes and decreases for transcriptionally upregulated genes. It is worth noting that the subtle ~5 – 10 bp difference in average linker length between highly-transcribed genes and poorly transcribed genes cannot simply explain the magnitude of differences in gene compaction changes, even though nucleosomes over highly transcribed genes are located closer to one another than nucleosomes over in poorly transcribed genes (WEINER *et al.* 2010). Although transcriptional activity appears to be the major driving force for chromatin folding at the gene level, this only accounts for ~31% of gene compaction, suggesting other “crumple factors” (e.g., chromatin remodelers, histone modifiers, architecture proteins, and

topoisomerases) may also contribute to the folding behavior of genes (**Figures 2.5A-C**). We investigated chromosome folding in detail in 14 mutants, confirming two categories of regulation in a transcription-dependent or -independent manner (**Figure 2.5D**). Increasing compaction in ribosomal protein genes and highly-transcribed genes dominated the signal in the transcription-dependent group including mutants in the RSC-SCC and Pol2-associated complexes. In contrast, mutations in chromatin regulators result in a global loss of gene compaction, suggesting a structural role in chromatin folding.

As mentioned above, in addition to their roles as domain insulators in yeast, the RSC complex and Cohesin loader promote transcriptional firing via maintenance of nucleosome depletion at promoters (HARTLEY AND MADHANI 2009; LOPEZ-SERRA *et al.* 2014), as well as facilitating Pol2 processing by evicting nucleosomes while traveling with the Pol2 machinery (SPAIN *et al.* 2014). Mutation of the RSC-SCC complex severely impaired transcriptional activity, resulting in gene condensing particularly in highly transcribed genes, consistent with the results of transcription perturbations (**Figures 2.4E-H** and **2.6C-D**). However, boundary calling is not accurate with such low-coverage datasets, leaving the functional consequences between boundaries and the RSC-SCC complex for future investigation. Moreover, the mutation of “gene looping factor” Ssu72 resulted in a moderate but significant loss of over 70% of gene crumples, and an increase in compaction for highly-transcribed genes, suggesting that Ssu72 may mediate chromatin compaction in both models (**Figures 2.6A-B**).

The multivalent Mediator complex roughly encompasses 25 subunits in yeast and over 30 subunits in mammals. It consists of five modules (Tail, Middle, Head, Scaffold, and Kinase), and a detailed structure been resolved recently by cryo-EM (TSAI *et al.* 2014; PLASCHKA *et al.* 2015). The Head, Middle, and Scaffold modules are essential for viability, while depletion of the Tail and Kinase modules only causes minor effects in yeast. Mediator complex has been seen as a transcriptional activator in earlier studies, with the capacity to promote pre-initiation (PIC) assembly, recruit Pol2, and stimulate the phosphorylation of the Pol2 CTD. Recent studies revealed Mediator can undergo compositional remodeling and participate many more biological processes than previously thought, such as transcription elongation, termination, mRNA splicing, DNA looping, chromatin remodeling, histone modification, and DNA methylation, as well mediating enhancer-promoter looping in metazoans (KAGEY *et al.* 2010; MALIK AND ROEDER 2010; NOCK *et al.* 2012; CARLSTEN *et al.* 2013; PHILLIPS-CREMINS *et al.* 2013; KEMMEREN *et al.* 2014; ALLEN AND TAATJES 2015). These complicated functions lead to controversial findings in Mediator studies, as even their chromatin binding profiles often disagree with each other. A consensus opinion that the Tail module is required for Mediator recruitment to the UAS, but dispensable to the core promoter regions; in other words, the Head and Middle modules can bind on the core promoter regions without need of the Tail, suggesting discrete functions (GRUNBERG *et al.* 2016; JERONIMO *et al.* 2016; PETRENKO *et al.* 2016). In addition, the mutations of 14 non-essential Mediator

subunits revealed the uncoupling of effects on transcription regulation, with changes in gene expression clustered by the groups of modules (VAN DE PEPPEL *et al.* 2005; KEMMEREN *et al.* 2014). Such evidence of multivalent Mediator modules explains the disparity of gene compaction effects between the mutants in the Tail and Middle (**Figures 2.6E-I**). We proposed that the Tail mainly functions on transcriptional activation via bridging the UAS-activator to the transcriptional machinery, and affects chromatin compaction in a transcription-dependent manner. On the contrary, the Head and Middle may play structural roles in chromatin folding regardless of transcriptional regulation, at least at the gene level.

Other candidates (**Figures 2.5D-I**) such as architectural proteins (Cohesin, Condensin) and histone modifiers (Rpd3, Swr1) regulate chromatin folding largely independent of transcriptional regulation, while the R-loop cleaner (Rnh201) or Topoisomerase (Top1) have subtle or target-specific effects on gene compaction. Interestingly, Rtt109, an H3K56 acetyltransferase which enhances replication-independent histone turnover, has global effects on chromosome compaction. The possibility that incorporating newly-synthesized histones influences chromatin folding opens an avenue for further exploration.

Several mutants in this study, including *ssu72-2*, *H4Δ4-14*, and *med1Δ*, exhibit dramatic decreases in nucleosome interactions as measured by Micro-C. In each of these cases, the loss of interactions is observed in normalized data, meaning that the “lost” interactions are compensated for by a gain in nonspecific

N/N+1 dinucleosome capture, and nonspecific long-range (N/N+60 or more) interactions. The simplest interpretation of a global loss of Micro-C interactions is as loss of physical interactions between nucleosomes, but more generally results from an inability to crosslink nucleosomes to one another, whether due to a lack of physical contact between nucleosomes, or obscuring of or loss of lysines involved in crosslinking. For some mutants – most notably *H4Δ4-14* – it is quite plausible that crosslinking artifacts cause the loss of Micro-C interactions, while this explanation is unlikely for mutants such as *ssu72-2*, *rtt109Δ*, and *med1Δ*.

Micro-C XL: improving capture of long-range interactions

While the improved resolution afforded by Micro-C enabled the identification of fine-scale features such as chromosomally-interacting domains (CIDs) in budding yeast that were not discernible in typical restriction-based 3C techniques, the known high-order chromosomal interactions such as centromere clusters were poorly recovered by Micro-C. In addition, as discussed above, we found no evidence for gene loops in Micro-C data, possibly due to using different fractions of digested chromatin or different levels of detergent treatment. These discrepancies with the literature motivated a deeper exploration of the effects of specific protocol steps on the results of Micro-C analysis of chromosome folding. In Chapter III, we developed an improved Micro-C protocol, termed Micro-C XL, which exhibits greatly increased signal to noise ratio, and provides further insight into the folding of the yeast genome.

Two modified steps greatly improved the Micro-C protocol in capturing higher-order interactions while maintaining the ability to map chromatin at single-nucleosome resolution. First, in addition to the “zero-length” crosslinker formaldehyde (3 Å), longer crosslinkers were supplemented to crosslink chromatin interactions. In our preliminary screening, we successfully identified two crosslinkers (DSG (7.7 Å) and EGS (16.1 Å)) that are capable of enhancing interactions between genomic loci at longer ranges of distance surrounding ~6 kb of MDJ1 regions and between centromeres (**Figures 3.1B-F**). Genome-wide interaction maps confirmed that the dual crosslinking protocol significantly improved the recovery of high-order chromatin features, with the identification of centromere clusters and Rabl-like chromosome structures (**Figure 3.2**). Second, fragmented chromatin was separated into soluble and insoluble fractions prior to ligation of crosslinked genomic loci. The signals of chromosome features, such as centromere and telomere clusters, were sharpened further with the addition of insoluble materials, compared to using only soluble material or combining the fractions for Micro-C, suggesting longer distance interactions (or overall chromatin interactions) are predominantly enriched in the insoluble over the soluble fraction (**Figure 3.3**). Recent technical reports for typical Hi-C methods agree with the interaction data from insoluble materials (e.g., *in pellet*, *in situ*, *in nucleus*), and yielded more consistent, reproducible, and cleaner datasets (NAGANO *et al.* 2015b).

Contrary to our initial expectations, the dramatic improvement seen in the apparent capture of long-range interactions using the Micro-C XL protocol likely results not from the ability of long-range crosslinkers to bridge interacting genomic loci associated with proteins that are more than 3 Å away from one another, but rather from a decrease in the noise caused by soluble nucleosomes encountering one another in solution during the ligation reaction, and causing artefactual byproducts between unlinked nucleosomes. This hypothesis is based on the fact that DSG- and EGS-based Micro-C maps are extremely similar despite their substantial difference in crosslinking distance, as well as the finding that isolation of soluble chromatin results in greatly increased noise in Micro-C maps. In addition, we note that chromatin fragments generated by restriction enzymes in typical Hi-C protocols are significantly larger than mononucleosomes, increasing the number of crosslinking opportunities per fragment and thus presumably restricting their diffusion (or solubility) and resultant ability to generate artefactual ligation products. We propose that this difference in fragment size/mobility accounts for the increased noise seen previously in Micro-C relative to standard Hi-C protocols. Further supporting this idea, we find that the improved protocol strongly reduces the incidence of artefactual ligation products between the nuclear genome and the mitochondrial genome, relative to the standard Micro-C protocol. Together, these considerations support the idea that the use of long crosslinkers and isolation of insoluble chromatin may be important to prevent mononucleosomes from freely diffusing prior to ligation and

introducing noise into Micro-C measurements. Still, this does not rule out the additional possibility that in some cases, our long crosslinkers capture nearby genomic loci for which the closest crosslinkable proteins are not in immediate physical proximity, and indeed both of these features may contribute to the improvement in data quality seen in Micro-C XL.

The improved technique Micro-C XL enables mapping of chromatin conformation at a full-scale range, from a single nucleosome to the entire genome, in a single assay, with the potential to be applied broadly to other species and organisms. A comparison of chromosome folding maps of *S. cerevisiae* and *S. pombe* by Micro-C XL revealed widespread qualitative similarities, yet quantitative differences, between these distantly-related species (**Figure 3.4**). At a high-order scale, Rabl-like chromosome conformation was clearly observed in both budding and fission yeasts, with multiple known characteristics seen in microscopic and Hi-C studies being recapitulated by Micro-C XL, including 1) centromere clustering, 2) avoidance of interaction between centromere and chromosome arms, and 3) telomere interactions of chromosome arms with similar length. Although the global chromosome organizations are very similar between the two species, we found *S. pombe* has stronger telomere-telomere interactions and a moderate increase in centromere clustering, leading to the presumption that the H3K9me3-mediated heterochromatin enhances the interactions between these repetitive regions in fission yeast. On a finer scale, individual genes tend to fold into crumples/globule

structures rather than loops in both species. Even the relatively insoluble materials were used in the protocol as prior gene loop studies, we found no evidence to support gene loop as a common structure existing *in vivo*. We were only able to identify on average five chromatin loops after screening through the entire yeast genome, with the sizes being much larger than the scale of a single gene. Chromatin domains (CIDs) broadly encompass ~1 – 5 genes, ~2 – 10 kb in both species, which can be further merged with neighboring domains in a hierarchical manner to form larger self-associating domains at a gradually increasing size (e.g., 20 kb, 40 kb, 80 kb). Boundaries between CIDs generally occur at active promoters, highly-transcribed genes, and tRNA genes, with the strongest group of boundaries localizing to divergent promoters or longer NDRs. In addition, the distinct slopes of interaction decay curves between species revealed a unique chromatin structure at size ~ 80 kb in *S. pombe*, reflecting the Cohesin-delimited globule found in a previous Hi-C study but not observed in budding yeast (MIZUGUCHI *et al.* 2014). Further analysis revealed that genomic regions surrounding Cohesin-associated loci are relatively insulated from physically interacting with one another. However, this insulation is stronger and persists over greater genomic distance in *S. pombe* relative to *S. cerevisiae*, hinting at important differences in the role of Cohesin between species, potentially in a cell-cycle dependent fashion. In conclusion, the modified protocol – Micro-C XL – is capable of mapping genome-wide, 3D chromatin structures at all scales of resolution in a single assay. The method is highly applicable to

different organisms such as drosophila and mammals, and broadly utilized in studying a magnitude of biological systems as discussed briefly below.

Determinants of chromatin folding in yeast

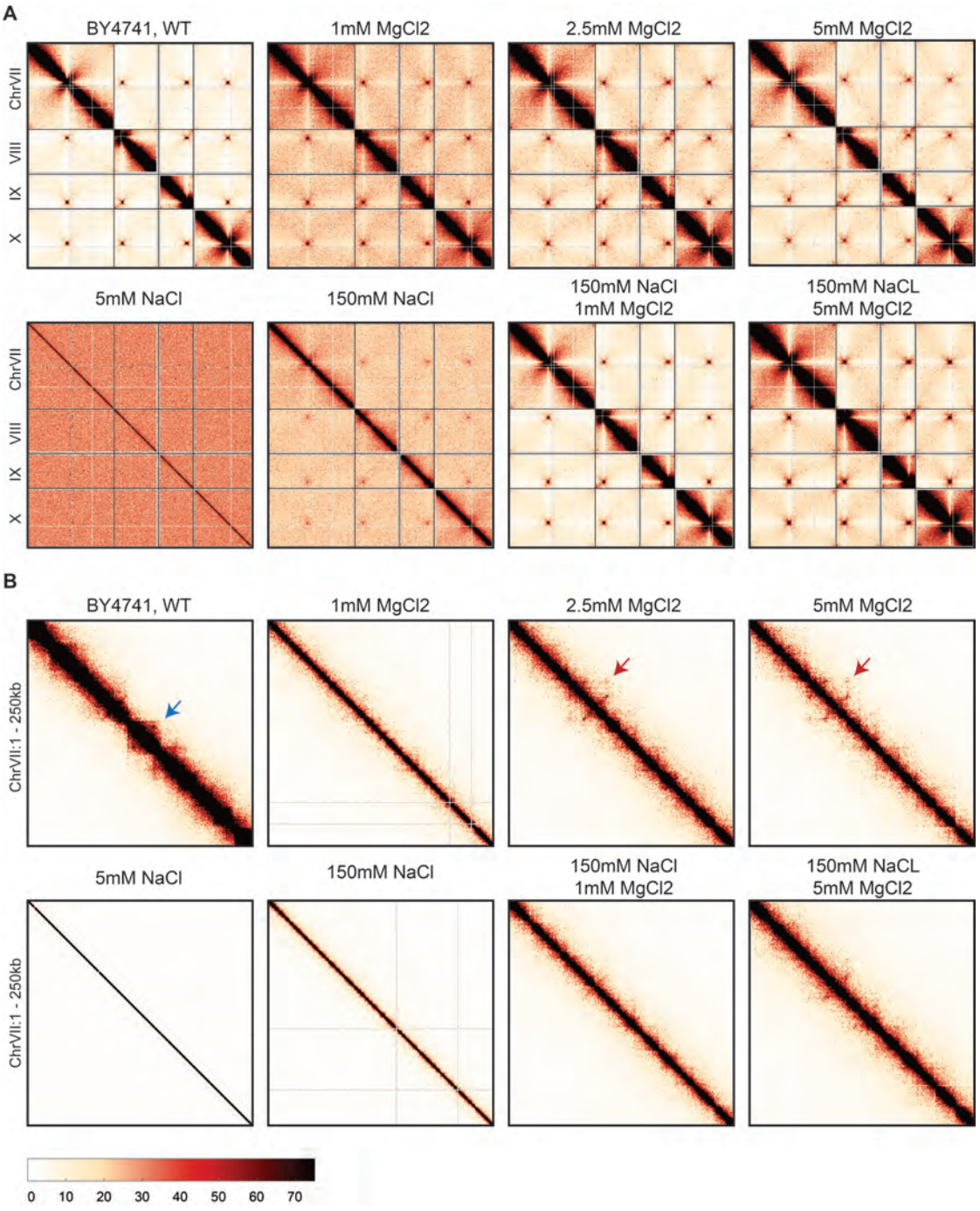
Biochemical and microscopic studies have identified many key factors that can influence chromatin folding behaviors such as salt concentrations, molecular crowding, the presence of histone H1, and nucleosome repeat lengths (LI AND REINBERG 2011; GRIGORYEV AND WOODCOCK 2012). However, most studies used identical repeats of the Widom 601 nucleosome positioning sequence, which lacks the key features of native chromatin (LOWARY AND WIDOM 1998). For example, nucleosome positions are not evenly spaced *in vivo*, genomic features such as poly A/T sequences can affect the stiffness and helical pitch of the DNA fiber, and histones are subject to a multitude of chemical modifications that essentially alter the properties of nucleosomes. We aim to dissect the role of individual factors in regulating chromatin folding by using a native chromatin template in both *in vitro* (bottom-up) and *ex vivo* (top-down) systems.

A key goal of the proposed research is to manipulate biochemical conditions coupled with genome-wide analysis of chromosome conformation (e.g., Micro-C) to understand the principles underlying the folding of the genome. In preliminary trials, we first established the top-down system in yeast that is based on the permeabilization of yeast cells to enable biochemical access to otherwise native chromatin. In brief, yeast cells were permeabilized by

treatments with enzyme and detergent, permeabilized nuclei were dialyzed *in situ* by eight different buffers, and *ex vivo* reconstituted nuclei were subjected to the Micro-C XL protocol and library construction. To this end, we have generated preliminary chromosome folding maps from permeabilized yeast nuclei subjected to a variety of salt conditions including 1) 1mM EDTA, 2) 1mM Mg^{2+}, 3) 2.5mM Mg^{2+}, 4) 5mM Mg^{2+}, 5) 5mM Na^{+} and 1mM EDTA, 6) 150mM Na^{+} and 1mM EDTA, 7) 150mM Na^{+} and 1mM Mg^{2+}, and 8) 150mM Na^{+} and 5 mM Mg^{2+}, expecting to fully or partially unfold native chromosomes.

The Micro-C data were broadly consistent with prior microscopic and biochemical studies (GRIGORYEV *et al.* 2009; GRIGORYEV AND WOODCOCK 2012) showing that monovalent and divalent cations can enhance compaction of nucleosome-repeat arrays *in vitro* or in permeabilized nuclei (**Figure 5.1**). We confirmed a nearly complete loss of chromatin interactions – observed as either the width of the diagonal in contact heatmaps, or as the slope of interaction frequency as a function of genomic distance – in yeast nuclei washed with no (1mM EDTA) and low salt (5mM NaCl) conditions, intermediate levels of folding in 150 mM NaCl along, and strong retention of native chromatin features in 150mM NaCl and 1mM $MgCl_2$, or in higher concentrations of $MgCl_2$ (**Figure 5.1A** and **C**). Interactions between interchromosomal arms, centromeres, and telomeres are further enhanced in the presence of divalent cations, indicating their global chromosomal conformations are more condensed than wild-type. Curiously, although we found that CEN-CEN interactions are maintained in salt-

washed nuclei in a Mg^{2+}-dependent manner, the centromere cluster was unexpectedly maintained in nuclei washed by 150mM NaCl. In addition, despite a high concentration of Mg^{2+}-induced overall chromosomal compaction, we found a multitude of local changes in chromatin including loss of a subset of boundaries and chromatin domains, and formation of “neo”-chromatin loops (**Figure 5.1B**). These preliminary results demonstrate that Micro-C can characterize the folding behavior of biochemically accessible chromatin from yeast, which in general recapitulates the features of chromatin folding observed in prior studies, while providing additional genomic information at single nucleosome resolution.


A
BY4741, WT
1mM MgCl2
2.5mM MgCl2
5mM MgCl2
ChrVII
VIII
IX
X
5mM NaCl
150mM NaCl
150mM NaCl
1mM MgCl2
150mM NaCL
5mM MgCl2
B
ChrVII:1 - 250kb
0 10 20 30 40 50 60 70

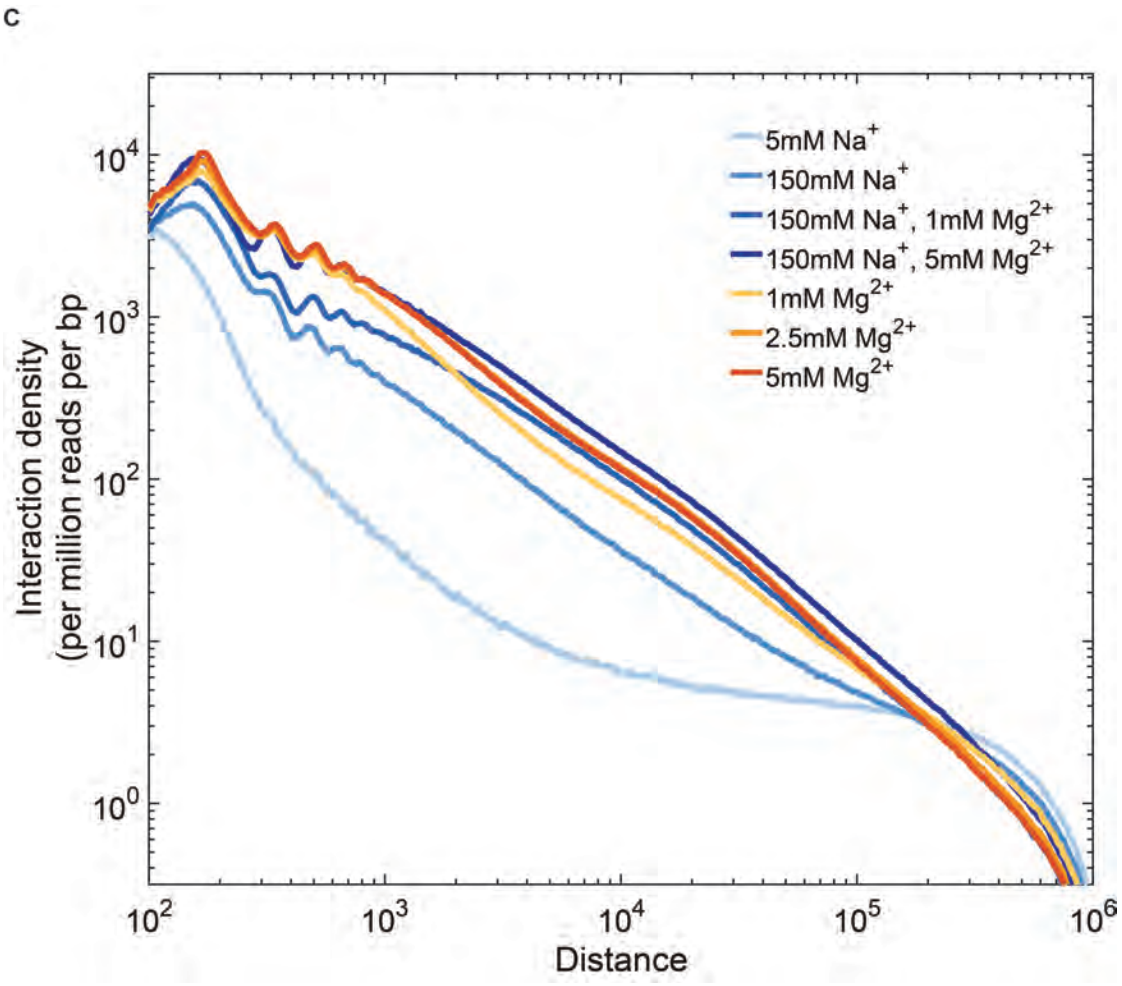


Figure 5.1. Salt-dependent chromatin folding.

(**A-B**) Contact maps for 7 of the salt conditions as indicated, compared to *in vivo* chromatin (BY4741, WT). In (A), the interaction maps shown for chromosome VII through X were binned to 5 kb and corrected by matrix balance. Global chromatin interactions are disrupted in low-salt condition (5mM NaCl) and moderately maintained in mid-range of salt conditions (150mM NaCl). The large-scale chromosomal interactions such as centromere and telomere clusters are maintained or even strengthened in the presence of divalent cation. More surprisingly, the intermediate level of salt condition (150mM NaCl) recovered the centromere interactions. In (B), the 1kb-binned interaction maps were corrected by matrix balance and further zoomed into a 250kb region on chromosome VII. The local chromatin folding is also largely dependent on salt conditions. Interestingly, some chromatin structures (e.g., domains or loops) were

disrupted in ex vivo conditions (blue arrow), while some neo chromatin structures were formed in the divalent salt buffers (red arrow). (**C**) Plots show chromatin interactions (y-axis, $\log_{10}$) vs. distance (x-axis, $\log_{10}$) decay curves for 7 of the salt reconstitutions. Interactions are globally lost in chromatin washed with low salt buffers, and partially maintained in 150mM NaCl alone, and are highly similar to native chromatin in the presence of divalent cation. Note that the distinct decaying behaviors through 1kb to 1Mb between different concentrations of Mg^{2+} (e.g., 1mM Mg^{2+} vs. 5mM Mg^{2+} vs. 150mM Na^{+}+5mM Mg^{2+}) suggest Mg^{2+}-dependent chromatin organizations beyond local short-range interactions.

Future research will move beyond cation concentration, and explore a wide range of additional conditions to check for effects on chromatin folding. This includes treating of nuclei with various enzymes, such as RNase H to remove R-loops, or light trypsin digestion to remove histone tails. Second, both our preliminary results and recent atomic force microscopic studies (KRZEMIEN *et al.* 2017) show that salt mediates chromatin compaction in a non-monotonic fashion including roughly three phases from low to high concentration, leading us to map chromatin conformations in a smaller range of salt condition (e.g. 5mM, 10mM, 20mM). In addition, very high salt conditions (> 500mM) strip off non-histone binding proteins, such as transcription factors or chromatin remodelers, from chromatin without affecting the binding between histones and DNA. This will allow us to compare chromatin folding behaviors with the presence or the absence of non-histone DNA binding proteins bound to chromatin. Finally, we will investigate the role of intrinsic forces in folding chromatin, measuring the reversibility of chromatin folding by carrying the successive washes in different

strengths of folding buffers. We showed that chromatin washed with a low salt buffer exhibits a widespread decompaction of chromatin, along with loss of centromere interactions. Does this loss of compaction result from key folding proteins being washed off the genome under low-salt conditions? Can interactions be restored upon reintroduction of Mg^{2+}? To this end, we will generate Micro-C datasets from permeabilized nuclei washed in low salt conditions, then washed a final time with high salt to “re-fold” the genome. The proposed research aims to promote our understanding of the mechanistic principles of chromatin folding with a genome-wide view.

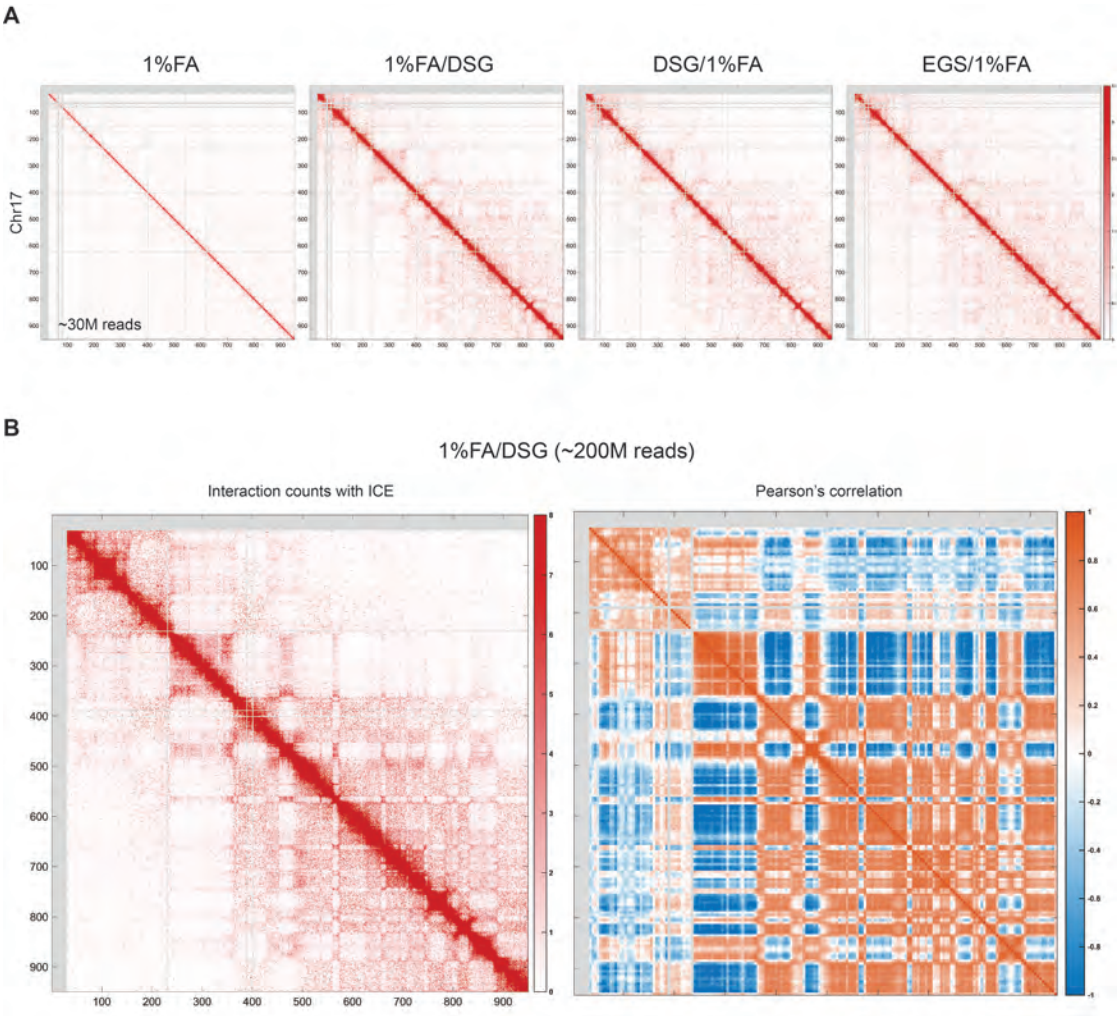
Toward a nucleosome-scale of chromatin structure in mammals

A major concern in applying Micro-C to large genome organisms (such as human and mouse) is the possibility that achieving sufficient sequencing coverage to visualization a single nucleosome level map may come at a high cost. As discussed in Chapter III, given that Micro-C data does not display a preponderance of molecular byproducts, the sequencing depth required to achieve a given genomic resolution should be similar to or less than the typical Hi-C protocol. In addition, a key analysis of chromatin folding behavior is the decay of interaction frequency as a function of genomic distance, which is an averaged measure across the entire genome and is thus extremely robust in undersequenced data. We anticipate that very low coverage (< 5 million reads) of

Micro-C data will allow a robust comparison of average chromatin folding for Polycomb-repressed genes or HP1-associated regions.

To this end, we developed a modified Micro-C protocol for mammalian cells and generated a proof-of-principle group of Micro-C data for mouse embryonic stem cells, with five different crosslinking conditions including 1% formaldehyde, 1% formaldehyde prior to DSG or EGS, and DSG or EGS prior to 1% formaldehyde (**Figure 5.2**). Strikingly, although these libraries were extremely undersequenced, with only ~30 million reads per sample (**Figure 5.2A**), and the FA1%+DSG library resulting in ~200 million reads (**Figures 5.2B-F**), Micro-C robustly captured known chromosome features such as chromosome compartment A/B, topologically-associating domains (TADs), and even CTCF/Cohesin-mediated chromatin loops and short-range promoter-enhancer loops (**Figures 5.2B-D**). The analysis of interaction versus decay curves revealed distinct chromatin folding behaviors in 15 chromatin states (BOGU *et al.* 2015) on the scale of 100 – 5000 bp (~1 – 25 nucleosome), with overall highly compact folding in H3K9me3-enriched regions and loose folding in active and regulatory regions (**Figures 5.2E-F**). Curiously, the annotated polycomb-repressed chromatin (H3K27me3) has relatively lower levels of compaction in comparison to heterochromatin and elongating chromatin, which may reflect the bivalent features in mES cells. We also examined short-range chromatin folding in LADs, and found an increase in internucleosomal interactions within LADs compared to the interactions in the Pol2-enriched domains (data not shown).

Taken together, these preliminary results confirmed that our Micro-C protocol is extremely robust and can be applied to other organisms without the limitation of sequencing coverage. We anticipate using Micro-C to uncover more details in chromatin folding with unprecedented sensitivity and resolution for mapping chromatin conformations.



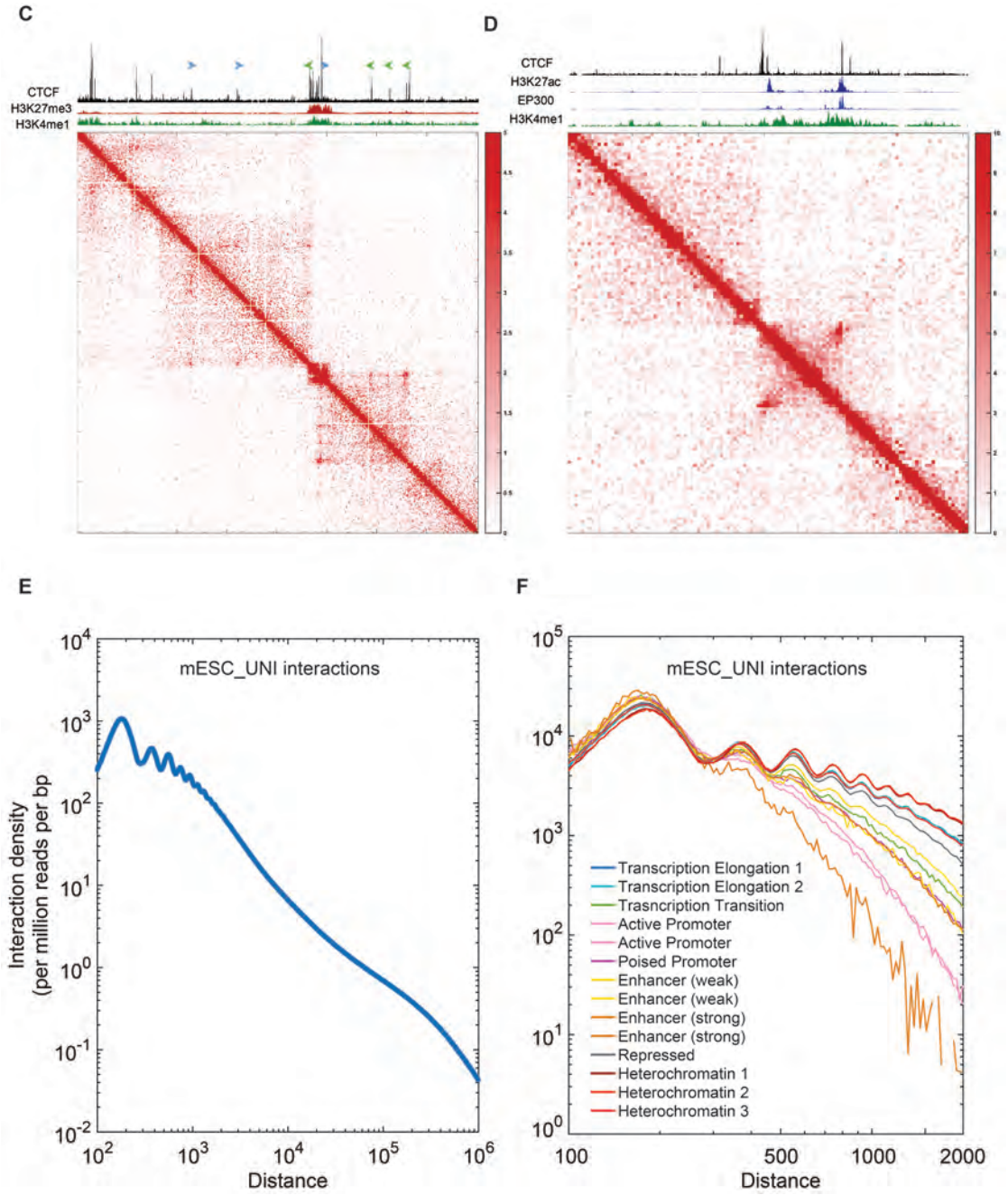


Figure 5.2. Mapping mammalian chromatin folding by Micro-C.

(**A**) Contact maps for Micro-C data generated using crosslinking conditions such as 1% Formaldehyde, 1% Formaldehyde following by DSG or EGS, and DSG or EGS following by 1% Formaldehyde are shown as indicated for chromosome 17 in mouse embryonic stem cells. The contact maps were binned to 500kb and normalized by matrix balancing

method. Although the preliminary datasets were only sequenced for ~30M reads per sample, it is worth noting that multiple-Mb chromosome compartments can be visualized as cells fixed with an additional crosslinker. (**B-F**) 1%FA/DSG library was further sequenced for ~200M reads. Interactions were normalized with sequence depth and corrected by matrix balance. In (B), the 500kb-binned contact matrix displays multiple-Mb chromosome compartments for chromosome 17 in mESC, and the compartmentalized genome (orange for compartment A and blue for compartment B) is clearer as shown in the Pearson's correlation map. In (C), the 5-kb binned interaction map was zoomed into 2Mb area surrounding the bivalent HoxD region. The bivalent regions separate two topologically-associating domains. Multiple CTCF-Cohesin mediated chromatin loops within an individual TAD are shown as sharp dots at the edge of TAD box. In (D), the sharp signal represents as the promoter-enhancer loop between Sox2 gene and 100kb downstream regulatory region. In (E), plot shows genome-wide averaged chromatin interactions (y-axis, $\log_{10}$) vs. distance (x-axis, $\log_{10}$) decay curve for mouse ES cells at the range from 100bp through 1Mb. The curve exhibits three stages of chromatin folding behavior including short-range nucleosome interactions/stacking (100bp – 1000bp), a monotonic decaying from 1kb – 100kb, and a bump representing chromatin domains around 100kb – 1Mb. In (F), plots show interactions vs. distance decay curves for 15 chromatin states at the scale of 100bp – 2000bp. Heterochromatic chromatin (H3K9me2/3) contains the highest level of local nucleosome interactions (or chromatin compaction), while a low level of nucleosomal interactions is found in the highly-transcribed and regulatory regions. Unexpectedly, the repressive chromatin enriched with H3K27me3 only has an intermediate level of compaction. Note that only the "UNI" interactions were used for the scaling plots in (E-F) in order to eliminate the undigested dinucleosomes.

Chromatin dynamics during a stress response

A major challenge in the chromatin field is understanding the role of chromatin marks in transcription. Multiple chromatin modifiers and readers form complexes that are recruited during transcription initiation and elongation, suggesting that

histone marks play a crucial role in transcriptional processes (SUGANUMA AND WORKMAN 2008; RANDO 2012). However, single gene and genome-wide studies have shown that genetic perturbation of such marks results in, at most, subtle differences in transcription (KORBER *et al.* 2006; WEINER *et al.* 2012). Moreover, previous studies in our lab showed that while loss of histone marks can be compensated for in steady state conditions, many effects of these mutants can be exposed during transcriptional reprogramming. In other words, because transcript levels are buffered by feedback mechanisms, many of them are restored to wild-type levels in deletion mutants (WEINER *et al.* 2012). This observation raised the questions of which chromatin marks change during such reprogramming and how these changes relate to the transcriptional machinery. These inspire us to further explore the effects of transcriptional reprogramming on histone modification dynamics.

In Chapter IV, we used multiplexed ChIP-seq to systematically map dynamic changes of 26 histone modifications in response to a stress signal during a dense time course in yeast. We generated a thorough characterization of the yeast chromatin landscape, which reproduces almost all known features of yeast histone modification localization (**Figures 4.1-4.3**), and provide further insights into the kinetics of histone modification (**Figures 4.4-4.7**). Analysis of histone modifications in actively growing yeast argues that histone marks exhibit little combinatorial complexity, as repeatedly observed in many studies (RANDO 2012), and conceptually disputes the histone code theory. In general, we

identified three major features that explain the patterns of nucleosome modification at steady state. First, the phosphorylation state delineates a specific subset of the primary chromosomal landscape; for example, pericentromeric regions are broadly marked by H3S10ph, and H2AS129ph marks subtelomeric and repressive regions. Second, the process of transcription leaves a massive footprint on chromatin. Enzymes associated with the initiation or elongation form of the Pol2 complex are responsible for the majority of the variation of histone modifications across the genome. Finally, replication-independent histone turnover is responsible for the deposition of histone marks such as H3K56ac, and lacks H3K79me3 and other marks. These features are well conserved in other organisms, although chromatin in many other organisms contains additional H3K9 and H3K27 di/trimethylations on repressive chromatin.

Analysis of a stress response revealed three key relationships observed between histone modifications and transcription reprogramming (**Figures 4.4-4.7**). First, the mechanisms that maintain the chromatin modification landscape in mid-log growth are similar to the ones involved in changes during stress-induced transcriptional reprogramming. In other words, modifications that are correlated with transcriptional activity in the steady state also increase during gene activation and decrease during gene repression, suggesting that the transcription machinery is responsible for depositing most transcription-related marks. Second, we only observed a modest increase in the combinatorial complexity of histone modification during diamide stress, resulting from roughly 3% of all

nucleosomes transiently populating rare histone modification states. Most of these rare histone states result from differences in the kinetics of histone modification that transiently uncouple highly correlated marks, with slow histone methylation changes often lagging behind more rapid acetylation changes. For example, H3K18 acetylation is usually deposited or removed ahead of H3K4 methylation in response to transcription activity. Finally, we carried out explicit analysis of histone modification dynamics, making a number of discoveries of interest that will motivate future studies. These include the general categorization of rapid and slow marks, the finding that modification dynamics are significantly different for addition vs. removal of marks, and the discovery of differences in modification cascades for different biological gene sets. Most notably, ribosomal-related genes exhibit the unique regulatory kinetics during transcriptional repression. Taken together, the study in Chapter IV provides a comprehensive view of the chromatin landscape and the kinetics involved during massive transcriptional reprogramming.

Bibliography

Achinger-Kawecka, J., and S. J. Clark, 2017 Disruption of the 3D cancer genome blueprint. Epigenomics 9**:** 47-55.

Adkins, M. W., J. J. Carson, C. M. English, C. J. Ramey and J. K. Tyler, 2007 The histone chaperone anti-silencing function 1 stimulates the acetylation of newly synthesized histone H3 in S-phase. J Biol Chem 282**:** 1334-1340.

Ahmed, K., H. Dehghani, P. Rugg-Gunn, E. Fussner, J. Rossant *et al.*, 2010 Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. PLoS One 5**:** e10531.

Akhtar, W., J. de Jong, A. V. Pindyurin, L. Pagie, W. Meuleman *et al.*, 2013 Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell 154**:** 914-927.

Al Husini, N., P. Kudla and A. Ansari, 2013 A role for CF1A 3' end processing complex in promoter-associated transcription. PLoS Genet 9**:** e1003722.

Albiez, H., M. Cremer, C. Tiberi, L. Vecchio, L. Schermelleh *et al.*, 2006 Chromatin domains and the interchromatin compartment form structurally defined and functionally interacting nuclear networks. Chromosome Res 14**:** 707-733.

Allen, B. L., and D. J. Taatjes, 2015 The Mediator complex: a central integrator of transcription. Nat Rev Mol Cell Biol 16**:** 155-166.

Ansari, A., and M. Hampsey, 2005 A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. Genes Dev 19**:** 2969-2978.

Arya, G., and T. Schlick, 2006 Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. Proc Natl Acad Sci U S A 103**:** 16236-16241.

Bau, D., A. Sanyal, B. R. Lajoie, E. Capriotti, M. Byron *et al.*, 2011 The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. Nat Struct Mol Biol 18**:** 107-114.

Beagan, J. A., T. G. Gilgenast, J. Kim, Z. Plona, H. K. Norton *et al.*, 2016 Local Genome Topology Can Exhibit an Incompletely Rewired 3D-Folding State during Somatic Cell Reprogramming. Cell Stem Cell 18**:** 611-624.

Bednar, J., R. A. Horowitz, S. A. Grigoryev, L. M. Carruthers, J. C. Hansen *et al.*, 1998 Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. Proc Natl Acad Sci U S A 95**:** 14173-14178.

Belton, J. M., B. R. Lajoie, S. Audibert, S. Cantaloube, I. Lassadi *et al.*, 2015 The Conformation of Yeast Chromosome III Is Mating Type Dependent and Controlled by the Recombination Enhancer. Cell Rep 13**:** 1855-1867.

Belton, J. M., R. P. McCord, J. H. Gibcus, N. Naumova, Y. Zhan *et al.*, 2012 Hi-C: a comprehensive technique to capture the conformation of genomes. Methods 58**:** 268-276.

Benedetti, F., J. Dorier, Y. Burnier and A. Stasiak, 2014 Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes. Nucleic Acids Res 42**:** 2848-2855.

Bernstein, B. E., T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert *et al.*, 2006 A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125**:** 315-326.

Blecher-Gonen, R., Z. Barnett-Itzhaki, D. Jaitin, D. Amann-Zalcenstein, D. Lara-Astiaso *et al.*, 2013 High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. Nat Protoc 8**:** 539-554.

Boettiger, A. N., B. Bintu, J. R. Moffitt, S. Wang, B. J. Beliveau *et al.*, 2016 Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature 529**:** 418-422.

Bogu, G. K., P. Vizan, L. W. Stanton, M. Beato, L. Di Croce *et al.*, 2015 Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse. Mol Cell Biol 36**:** 809-819.

Bonev, B., and G. Cavalli, 2016 Organization and function of the 3D genome. Nat Rev Genet 17**:** 661-678.

Boyle, A. P., S. Davis, H. P. Shulha, P. Meltzer, E. H. Margulies *et al.*, 2008 High-resolution mapping and characterization of open chromatin across the genome. Cell 132**:** 311-322.

Brant, L., T. Georgomanolis, M. Nikolic, C. A. Brackley, P. Kolovos *et al.*, 2016 Exploiting native forces to capture chromosome conformation in mammalian cell nuclei. Mol Syst Biol 12**:** 891.

Brogaard, K., L. Xi, J. P. Wang and J. Widom, 2012 A map of nucleosome positions in yeast at base-pair resolution. Nature 486**:** 496-501.

Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang and W. J. Greenleaf, 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.

Buratowski, S., 2009 Progression through the RNA polymerase II CTD cycle. Mol Cell 36**:** 541-546.

Carlsten, J. O., X. Zhu and C. M. Gustafsson, 2013 The multitalented Mediator complex. Trends Biochem Sci 38**:** 531-537.

Celic, I., H. Masumoto, W. P. Griffith, P. Meluh, R. J. Cotter *et al.*, 2006 The sirtuins hst3 and Hst4p preserve genome integrity by controlling histone h3 lysine 56 deacetylation. Curr Biol 16**:** 1280-1289.

Chandra, T., P. A. Ewels, S. Schoenfelder, M. Furlan-Magaril, S. W. Wingett *et al.*, 2015 Global reorganization of the nuclear landscape in senescent cells. Cell Rep 10**:** 471-483.

Chen, C., H. H. Lim, J. Shi, S. Tamura, K. Maeshima *et al.*, 2016 Budding yeast chromatin is dispersed in a crowded nucleoplasm in vivo. Mol Biol Cell 27**:** 3357-3368.

Cheng, T. M., S. Heeger, R. A. Chaleil, N. Matthews, A. Stewart *et al.*, 2015 A simple biophysical model emulates budding yeast chromosome condensation. Elife 4**:** e05565.

Churchman, L. S., and J. S. Weissman, 2011 Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469**:** 368-373.

Ciosk, R., M. Shirayama, A. Shevchenko, T. Tanaka, A. Toth *et al.*, 2000 Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol Cell 5**:** 243-254.

Clapier, C. R., and B. R. Cairns, 2009 The biology of chromatin remodeling complexes. Annu Rev Biochem 78**:** 273-304.

Collepardo-Guevara, R., and T. Schlick, 2012 Crucial role of dynamic linker histone binding and divalent ions for DNA accessibility and gene regulation revealed by mesoscale modeling of oligonucleosomes. Nucleic Acids Res 40**:** 8803-8817.

Collepardo-Guevara, R., and T. Schlick, 2014 Chromatin fiber polymorphism triggered by variations of DNA linker lengths. Proc Natl Acad Sci U S A 111**:** 8061-8066.

Crane, E., Q. Bian, R. P. McCord, B. R. Lajoie, B. S. Wheeler *et al.*, 2015 Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature 523**:** 240-244.

Crawford, G. E., S. Davis, P. C. Scacheri, G. Renaud, M. J. Halawi *et al.*, 2006a DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. Nat Methods 3**:** 503-509.

Crawford, G. E., I. E. Holt, J. Whittle, B. D. Webb, D. Tai *et al.*, 2006b Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res 16**:** 123-131.

Cremer, T., and M. Cremer, 2010 Chromosome territories. Cold Spring Harb Perspect Biol 2**:** a003889.

Criscione, S. W., M. De Cecco, B. Siranosian, Y. Zhang, J. A. Kreiling *et al.*, 2016a Reorganization of chromosome architecture in replicative cellular senescence. Sci Adv 2**:** e1500882.

Criscione, S. W., Y. V. Teo and N. Neretti, 2016b The Chromatin Landscape of Cellular Senescence. Trends Genet 32**:** 751-761.
Crosio, C., G. M. Fimia, R. Loury, M. Kimura, Y. Okano *et al.*, 2002 Mitotic phosphorylation of histone H3: spatio-temporal regulation by mammalian Aurora kinases. Mol Cell Biol 22**:** 874-885.
Cui, Y., and C. Bustamante, 2000 Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure. Proc Natl Acad Sci U S A 97**:** 127-132.
Dai, J., E. M. Hyland, D. S. Yuan, H. Huang, J. S. Bader *et al.*, 2008 Probing nucleosome function: a highly versatile library of synthetic histone H3 and H4 mutants. Cell 134**:** 1066-1078.
De Vos, D., F. Frederiks, M. Terweij, T. van Welsem, K. F. Verzijlbergen *et al.*, 2011 Progressive methylation of ageing histones by Dot1 functions as a timer. EMBO Rep 12**:** 956-962.
de Wit, E., E. S. Vos, S. J. Holwerda, C. Valdes-Quezada, M. J. Verstegen *et al.*, 2015 CTCF Binding Polarity Determines Chromatin Looping. Mol Cell 60**:** 676-684.
Deal, R. B., and S. Henikoff, 2010 Capturing the dynamic epigenome. Genome Biol 11**:** 218.
Dekker, J., 2008 Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction. J Biol Chem 283**:** 34532-34540.
Dekker, J., M. A. Marti-Renom and L. A. Mirny, 2013 Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet 14**:** 390-403.
Dekker, J., and T. Misteli, 2015 Long-Range Chromatin Interactions. Cold Spring Harb Perspect Biol 7**:** a019356.
Dekker, J., K. Rippe, M. Dekker and N. Kleckner, 2002 Capturing chromosome conformation. Science 295**:** 1306-1311.
DeMare, L. E., J. Leng, J. Cotney, S. K. Reilly, J. Yin *et al.*, 2013 The genomic landscape of cohesin-associated chromatin interactions. Genome Res 23**:** 1224-1234.
Denker, A., and W. de Laat, 2016 The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev 30**:** 1357-1382.
Dion, M. F., T. Kaplan, M. Kim, S. Buratowski, N. Friedman *et al.*, 2007 Dynamics of replication-independent histone turnover in budding yeast. Science 315**:** 1405-1408.
Dixon, J. R., I. Jung, S. Selvaraj, Y. Shen, J. E. Antosiewicz-Bourget *et al.*, 2015 Chromatin architecture reorganization during stem cell differentiation. Nature 518**:** 331-336.
Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li *et al.*, 2012 Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485**:** 376-380.
Dorigo, B., T. Schalch, A. Kulangara, S. Duda, R. R. Schroeder *et al.*, 2004 Nucleosome arrays reveal the two-start organization of the chromatin fiber. Science 306**:** 1571-1573.
Dostie, J., and J. Dekker, 2007 Mapping networks of physical interactions between genomic elements using 5C technology. Nat Protoc 2**:** 988-1002.
Dostie, J., T. A. Richmond, R. A. Arnaout, R. R. Selzer, W. L. Lee *et al.*, 2006 Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res 16**:** 1299-1309.
Dowen, J. M., Z. P. Fan, D. Hnisz, G. Ren, B. J. Abraham *et al.*, 2014 Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell 159**:** 374-387.
Doyle, B., G. Fudenberg, M. Imakaev and L. A. Mirny, 2014 Chromatin loops as allosteric modulators of enhancer-promoter interactions. PLoS Comput Biol 10**:** e1003867.
Duan, Z., M. Andronescu, K. Schutz, S. McIlwain, Y. J. Kim *et al.*, 2010 A three-dimensional model of the yeast genome. Nature 465**:** 363-367.
Eagen, K. P., T. A. Hartl and R. D. Kornberg, 2015 Stable Chromosome Condensation Revealed by Chromosome Conformation Capture. Cell 163**:** 934-946.
Eagen, K. P., E. Lieberman Aiden and R. D. Kornberg, 2017 Polycomb-Mediated Chromatin Loops Revealed by a Sub-Kilobase Resolution Chromatin Interaction Map. bioRxiv.

El Kaderi, B., S. Medler, S. Raghunayakula and A. Ansari, 2009 Gene looping is conferred by activator-dependent interaction of transcription initiation and termination machineries. J Biol Chem 284**:** 25015-25025.

Eltsov, M., K. M. Maclellan, K. Maeshima, A. S. Frangakis and J. Dubochet, 2008 Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. Proc Natl Acad Sci U S A 105**:** 19732-19737.

Enervald, E., E. Lindgren, Y. Katou, K. Shirahige and L. Strom, 2013 Importance of Poleta for damage-induced cohesion reveals differential regulation of cohesion establishment at the break site and genome-wide. PLoS Genet 9**:** e1003158.

Ernst, J., P. Kheradpour, T. S. Mikkelsen, N. Shoresh, L. D. Ward *et al.*, 2011 Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473**:** 43-49.

Fan, X., D. M. Chou and K. Struhl, 2006 Activator-specific recruitment of Mediator in vivo. Nat Struct Mol Biol 13**:** 117-120.

Felsenfeld, G., and M. Groudine, 2003 Controlling the double helix. Nature 421**:** 448-453.

Feng, S., S. J. Cokus, V. Schubert, J. Zhai, M. Pellegrini *et al.*, 2014 Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis. Mol Cell 55**:** 694-707.

Filion, G. J., J. G. van Bemmel, U. Braunschweig, W. Talhout, J. Kind *et al.*, 2010 Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell 143**:** 212-224.

Finch, J. T., and A. Klug, 1976 Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A 73**:** 1897-1901.

Franke, M., D. M. Ibrahim, G. Andrey, W. Schwarzer, V. Heinrich *et al.*, 2016 Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature 538**:** 265-269.

Friedman, N., and O. J. Rando, 2015 Epigenomics and the structure of the living genome. Genome Res 25**:** 1482-1490.

Fudenberg, G., M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur *et al.*, 2016 Formation of Chromosomal Domains by Loop Extrusion. Cell Rep 15**:** 2038-2049.

Fullwood, M. J., M. H. Liu, Y. F. Pan, J. Liu, H. Xu *et al.*, 2009 An oestrogen-receptor-alpha-bound human chromatin interactome. Nature 462**:** 58-64.

Fussner, E., R. W. Ching and D. P. Bazett-Jones, 2011 Living without 30nm chromatin fibers. Trends Biochem Sci 36**:** 1-6.

Fussner, E., M. Strauss, U. Djuric, R. Li, K. Ahmed *et al.*, 2012 Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. EMBO Rep 13**:** 992-996.

Gall, J., 1963 Chromosome fibers from an interphase nucleus. Science 139**:** 120-121.

Gall, J. G., 1966 Chromosome fibers studied by a spreading technique. Chromosoma 20**:** 221-233.

Garber, M., N. Yosef, A. Goren, R. Raychowdhury, A. Thielke *et al.*, 2012 A High-Throughput Chromatin Immunoprecipitation Approach Reveals Principles of Dynamic Gene Regulation in Mammals. Mol Cell 47**:** 810-822.

Gasch, A. P., P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen *et al.*, 2000 Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell 11**:** 4241-4257.

Gavrilov, A. A., E. S. Gushchanskaya, O. Strelkova, O. Zhironkina, Kireev, II *et al.*, 2013 Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub. Nucleic Acids Res 41**:** 3563-3575.

Gheldof, N., T. M. Tabuchi and J. Dekker, 2006 The active FMR1 promoter is associated with a large domain of altered chromatin conformation with embedded local histone modifications. Proc Natl Acad Sci U S A 103**:** 12463-12468.

Ghirlando, R., and G. Felsenfeld, 2008 Hydrodynamic studies on defined heterochromatin fragments support a 30-nm fiber having six nucleosomes per turn. J Mol Biol 376**:** 1417-1425.

Gibcus, J. H., and J. Dekker, 2013 The hierarchy of the 3D genome. Mol Cell 49**:** 773-782.
Gierman, H. J., M. H. Indemans, J. Koster, S. Goetze, J. Seppen *et al.*, 2007 Domain-wide regulation of gene expression in the human genome. Genome Res 17**:** 1286-1295.
Gilbert, N., S. Boyle, H. Fiegler, K. Woodfine, N. P. Carter *et al.*, 2004 Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. Cell 118**:** 555-566.
Grigoryev, S. A., G. Arya, S. Correll, C. L. Woodcock and T. Schlick, 2009 Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. Proc Natl Acad Sci U S A 106**:** 13317-13322.
Grigoryev, S. A., G. Bascom, J. M. Buckwalter, M. B. Schubert, C. L. Woodcock *et al.*, 2016 Hierarchical looping of zigzag nucleosome chains in metaphase chromosomes. Proc Natl Acad Sci U S A 113**:** 1238-1243.
Grigoryev, S. A., and C. L. Woodcock, 2012 Chromatin organization - the 30 nm fiber. Exp Cell Res 318**:** 1448-1455.
Grob, S., M. W. Schmid and U. Grossniklaus, 2014 Hi-C analysis in Arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila. Mol Cell 55**:** 678-693.
Groschel, S., M. A. Sanders, R. Hoogenboezem, E. de Wit, B. A. Bouwman *et al.*, 2014 A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell 157**:** 369-381.
Grunberg, S., S. Henikoff, S. Hahn and G. E. Zentner, 2016 Mediator binding to UASs is broadly uncoupled from transcription and cooperative with TFIID recruitment to promoters. EMBO J 35**:** 2435-2446.
Gruss, C., J. Wu, T. Koller and J. M. Sogo, 1993 Disruption of the nucleosomes at the replication fork. Embo J 12**:** 4533-4545.
Guelen, L., L. Pagie, E. Brasset, W. Meuleman, M. B. Faza *et al.*, 2008 Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 453**:** 948-951.
Guidi, M., M. Ruault, M. Marbouty, I. Loiodice, A. Cournac *et al.*, 2015 Spatial reorganization of telomeres in long-lived quiescent cells. Genome Biol 16**:** 206.
Guillemette, B., A. R. Bataille, N. Gevry, M. Adam, M. Blanchette *et al.*, 2005 Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. PLoS Biol 3**:** e384.
Guillemette, B., P. Drogaris, H. H. Lin, H. Armstrong, K. Hiragami-Hamada *et al.*, 2011 H3 lysine 4 is acetylated at active gene promoters and is regulated by H3 lysine 4 methylation. PLoS Genet 7**:** e1001354.
Guo, Y., Q. Xu, D. Canzio, J. Shou, J. Li *et al.*, 2015 CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell 162**:** 900-910.
Guttman, M., I. Amit, M. Garber, C. French, M. F. Lin *et al.*, 2009 Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458**:** 223-227.
Handoko, L., H. Xu, G. Li, C. Y. Ngan, E. Chew *et al.*, 2011 CTCF-mediated functional chromatin interactome in pluripotent cells. Nat Genet 43**:** 630-638.
Hartley, P. D., and H. D. Madhani, 2009 Mechanisms that specify promoter nucleosome location and identity. Cell 137**:** 445-458.
Heintzman, N. D., R. K. Stuart, G. Hon, Y. Fu, C. W. Ching *et al.*, 2007 Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39**:** 311-318.
Henikoff, J. G., J. A. Belsky, K. Krassovsky, D. M. MacAlpine and S. Henikoff, 2011 Epigenome characterization at single base-pair resolution. Proc Natl Acad Sci U S A 108**:** 18318-18323.
Henikoff, S., and K. Ahmad, 2005 Assembly of variant histones into chromatin. Annu Rev Cell Dev Biol 21**:** 133-153.

Henikoff, S., J. G. Henikoff, A. Sakai, G. B. Loeb and K. Ahmad, 2009 Genome-wide profiling of salt fractions maps physical properties of chromatin. Genome Res 19**:** 460-469.

Hizume, K., T. Nakai, S. Araki, E. Prieto, K. Yoshikawa *et al.*, 2009 Removal of histone tails from nucleosome dissects the physical mechanisms of salt-induced aggregation, linker histone H1-induced compaction, and 30-nm fiber formation of the nucleosome array. Ultramicroscopy 109**:** 868-873.

Hon, G., W. Wang and B. Ren, 2009 Discovery and annotation of functional chromatin signatures in the human genome. PLoS Comput Biol 5**:** e1000566.

Horn, P. J., and C. L. Peterson, 2002 Molecular biology. Chromatin higher order folding--wrapping up transcription. Science 297**:** 1824-1827.

Horowitz, R. A., D. A. Agard, J. W. Sedat and C. L. Woodcock, 1994 The three-dimensional architecture of chromatin in situ: electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon. J Cell Biol 125**:** 1-10.

Hou, C., L. Li, Z. S. Qin and V. G. Corces, 2012 Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Mol Cell 48**:** 471-484.

Hsieh, T. H., A. Weiner, B. Lajoie, J. Dekker, N. Friedman *et al.*, 2015 Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. Cell 162**:** 108-119.

Hu, B., N. Petela, A. Kurze, K. L. Chan, C. Chapard *et al.*, 2015 Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating ChIP-seq. Nucleic Acids Res 43**:** e132.

Hu, Y., I. Kireev, M. Plutz, N. Ashourian and A. S. Belmont, 2009 Large-scale chromatin structure of inducible genes: transcription on a condensed, linear template. J Cell Biol 185**:** 87-100.

Hubner, B., M. Lomiento, F. Mammoli, D. Illner, Y. Markaki *et al.*, 2015 Remodeling of nuclear landscapes during human myelopoietic cell differentiation maintains co-aligned active and inactive nuclear compartments. Epigenetics Chromatin 8**:** 47.

Hughes, A. L., and O. J. Rando, 2014 Mechanisms underlying nucleosome positioning in vivo. Annu Rev Biophys 43**:** 41-63.

Hughes, J. R., N. Roberts, S. McGowan, D. Hay, E. Giannoulatou *et al.*, 2014 Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat Genet 46**:** 205-212.

Imakaev, M., G. Fudenberg, R. P. McCord, N. Naumova, A. Goloborodko *et al.*, 2012 Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods 9**:** 999-1003.

Imakaev, M. V., G. Fudenberg and L. A. Mirny, 2015 Modeling chromosomes: Beyond pretty pictures. FEBS Lett 589**:** 3031-3036.

Jansen, A., and K. J. Verstrepen, 2011 Nucleosome positioning in Saccharomyces cerevisiae. Microbiol Mol Biol Rev 75**:** 301-320.

Jenuwein, T., and C. D. Allis, 2001 Translating the histone code. Science 293**:** 1074-1080.

Jeronimo, C., M. F. Langelier, A. R. Bataille, J. M. Pascal, B. F. Pugh *et al.*, 2016 Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo. Mol Cell 64**:** 455-466.

Jiang, C., and B. F. Pugh, 2009 Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet 10**:** 161-172.

Joti, Y., T. Hikima, Y. Nishino, F. Kamada, S. Hihara *et al.*, 2012 Chromosomes without a 30-nm chromatin fiber. Nucleus 3**:** 404-410.

Kagey, M. H., J. J. Newman, S. Bilodeau, Y. Zhan, D. A. Orlando *et al.*, 2010 Mediator and cohesin connect gene expression and chromatin architecture. Nature 467**:** 430-435.

Kalhor, R., H. Tjong, N. Jayathilaka, F. Alber and L. Chen, 2011 Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nat Biotechnol 30**:** 90-98.

Kaplan, N., I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, D. Tillo *et al.*, 2009 The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 458**:** 362-366.

Kaplan, T., C. L. Liu, J. A. Erkmann, J. Holik, M. Grunstein *et al.*, 2008 Cell cycle- and chaperone-mediated regulation of H3K56ac incorporation in yeast. PLoS Genet 4**:** e1000270.

Keene, M. A., V. Corces, K. Lowenhaupt and S. C. Elgin, 1981 DNase I hypersensitive sites in Drosophila chromatin occur at the 5' ends of regions of transcription. Proc Natl Acad Sci U S A 78**:** 143-146.

Kelly, T. K., Y. Liu, F. D. Lay, G. Liang, B. P. Berman *et al.*, 2012 Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res 22**:** 2497-2506.

Kemmeren, P., K. Sameith, L. A. van de Pasch, J. J. Benschop, T. L. Lenstra *et al.*, 2014 Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. Cell 157**:** 740-752.

Kent, N. A., S. Adams, A. Moorhouse and K. Paszkiewicz, 2011 Chromatin particle spectrum analysis: a method for comparative chromatin structure analysis using paired-end mode next-generation DNA sequencing. Nucleic Acids Res 39**:** e26.

Kim, T. S., C. L. Liu, M. Yassour, J. Holik, N. Friedman *et al.*, 2010 RNA polymerase mapping during stress responses reveals widespread nonproductive transcription in yeast. Genome Biol 11**:** R75.

Kind, J., L. Pagie, S. S. de Vries, L. Nahidiazar, S. S. Dey *et al.*, 2015 Genome-wide maps of nuclear lamina interactions in single human cells. Cell 163**:** 134-147.

Kind, J., L. Pagie, H. Ortabozkoyun, S. Boyle, S. S. de Vries *et al.*, 2013 Single-cell dynamics of genome-nuclear lamina interactions. Cell 153**:** 178-192.

Kizilyaprak, C., D. Spehner, D. Devys and P. Schultz, 2010 In vivo chromatin organization of mouse rod photoreceptors correlates with histone modifications. PLoS One 5**:** e11039.

Korber, P., S. Barbaric, T. Luckenbach, A. Schmid, U. J. Schermer *et al.*, 2006 The histone chaperone Asf1 increases the rate of histone eviction at the yeast PHO5 and PHO8 promoters. Journal of Biological Chemistry 281**:** 5539-5545.

Korber, P., and W. Horz, 2004 In vitro assembly of the characteristic chromatin organization at the yeast PHO5 promoter by a replication-independent extract system. J Biol Chem 279**:** 35113-35120.

Kornberg, R. D., 1974 Chromatin structure: a repeating unit of histones and DNA. Science 184**:** 868-871.

Kornberg, R. D., and Y. Lorch, 1999 Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell 98**:** 285-294.

Kornberg, R. D., and L. Stryer, 1988 Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. Nucleic Acids Res 16**:** 6677-6690.

Krietenstein, N., M. Wal, S. Watanabe, B. Park, C. L. Peterson *et al.*, 2016 Genomic Nucleosome Organization Reconstituted with Pure Proteins. Cell 167**:** 709-721 e712.

Krijger, P. H., and W. de Laat, 2016 Regulation of disease-associated gene expression in the 3D genome. Nat Rev Mol Cell Biol 17**:** 771-782.

Krijger, P. H., B. Di Stefano, E. de Wit, F. Limone, C. van Oevelen *et al.*, 2016 Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic Cell Reprogramming. Cell Stem Cell 18**:** 597-610.

Kruithof, M., F. T. Chien, A. Routh, C. Logie, D. Rhodes *et al.*, 2009 Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. Nat Struct Mol Biol 16**:** 534-540.

Krzemien, K. M., M. Beckers, S. Quack and J. Michaelis, 2017 Atomic force microscopy of chromatin arrays reveal non-monotonic salt dependence of array compaction in solution. PLoS One 12**:** e0173459.

Kubo, N., H. Ishii, D. Gorkin, F. Meitinger, X. Xiong *et al.*, 2017 Preservation of Chromatin Organization after Acute Loss of CTCF in Mouse Embryonic Stem Cells. bioRxiv.

Lajoie, B. R., J. Dekker and N. Kaplan, 2015 The Hitchhiker's guide to Hi-C analysis: practical guidelines. Methods 72**:** 65-75.

Le Dily, F., D. Bau, A. Pohl, G. P. Vicent, F. Serra *et al.*, 2014 Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. Genes Dev 28**:** 2151-2162.

Le, T. B., M. V. Imakaev, L. A. Mirny and M. T. Laub, 2013 High-resolution mapping of the spatial organization of a bacterial chromosome. Science 342**:** 731-734.

Le, T. B., and M. T. Laub, 2016 Transcription rate and transcript length drive formation of chromosomal interaction domain boundaries. EMBO J 35**:** 1582-1595.

Lee, K., C. C. Hsiung, P. Huang, A. Raj and G. A. Blobel, 2015 Dynamic enhancer-gene body contacts during transcription elongation. Genes Dev 29**:** 1992-1997.

Lenstra, T. L., J. J. Benschop, T. Kim, J. M. Schulze, N. A. Brabers *et al.*, 2011 The specificity and topology of chromatin interaction pathways in yeast. Mol Cell 42**:** 536-549.

Leuba, S. H., G. Yang, C. Robert, B. Samori, K. van Holde *et al.*, 1994 Three-dimensional structure of extended chromatin fibers as revealed by tapping-mode scanning force microscopy. Proc Natl Acad Sci U S A 91**:** 11621-11625.

Li, B., M. Carey and J. L. Workman, 2007 The role of chromatin during transcription. Cell 128**:** 707-719.

Li, G., and D. Reinberg, 2011 Chromatin higher-order structures and gene regulation. Curr Opin Genet Dev 21**:** 175-186.

Li, G., X. Ruan, R. K. Auerbach, K. S. Sandhu, M. Zheng *et al.*, 2012 Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 148**:** 84-98.

Li, L., X. Lyu, C. Hou, N. Takenaka, H. Q. Nguyen *et al.*, 2015 Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. Mol Cell 58**:** 216-231.

Liang, J., L. Lacroix, A. Gamot, S. Cuddapah, S. Queille *et al.*, 2014 Chromatin immunoprecipitation indirect peaks highlight long-range interactions of insulator proteins and Pol II pausing. Mol Cell 53**:** 672-681.

Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy *et al.*, 2009 Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326**:** 289-293.

Liu, C., C. Wang, G. Wang, C. Becker, M. Zaidem *et al.*, 2016 Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution. Genome Res 26**:** 1057-1068.

Liu, C. L., T. Kaplan, M. Kim, S. Buratowski, S. L. Schreiber *et al.*, 2005 Single-nucleosome mapping of histone modifications in S. cerevisiae. PLoS Biol 3**:** e328.

Liu, Z., and L. C. Myers, 2012 Med5(Nut1) and Med17(Srb4) are direct targets of mediator histone H4 tail interactions. PLoS One 7**:** e38416.

Lopez-Serra, L., G. Kelly, H. Patel, A. Stewart and F. Uhlmann, 2014 The Scc2-Scc4 complex acts in sister chromatid cohesion and transcriptional regulation by maintaining nucleosome-free regions. Nat Genet 46**:** 1147-1151.

Lowary, P. T., and J. Widom, 1998 New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol 276**:** 19-42.

Luger, K., M. L. Dechassa and D. J. Tremethick, 2012 New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nat Rev Mol Cell Biol 13**:** 436-447.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond, 1997 Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389**:** 251-260.

Lupianez, D. G., K. Kraft, V. Heinrich, P. Krawitz, F. Brancati *et al.*, 2015 Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161**:** 1012-1025.

Maas, N. L., K. M. Miller, L. G. DeFazio and D. P. Toczyski, 2006 Cell cycle and checkpoint regulation of histone H3 K56 acetylation by Hst3 and Hst4. Mol Cell 23**:** 109-119.

Maeshima, K., S. Hihara and M. Eltsov, 2010a Chromatin structure: does the 30-nm fibre exist in vivo? Curr Opin Cell Biol 22**:** 291-297.
Maeshima, K., S. Hihara and H. Takata, 2010b New insight into the mitotic chromosome structure: irregular folding of nucleosome fibers without 30-nm chromatin structure. Cold Spring Harb Symp Quant Biol 75**:** 439-444.
Maeshima, K., S. Ide, K. Hibino and M. Sasai, 2016 Liquid-like behavior of chromatin. Curr Opin Genet Dev 37**:** 36-45.
Maeshima, K., R. Imai, S. Tamura and T. Nozaki, 2014 Chromatin as dynamic 10-nm fibers. Chromosoma 123**:** 225-237.
Malik, S., and R. G. Roeder, 2010 The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet 11**:** 761-772.
Marbouty, M., A. Cournac, J. F. Flot, H. Marie-Nelly, J. Mozziconacci *et al.*, 2014 Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. Elife 3**:** e03318.
Martin, P., A. McGovern, J. Massey, S. Schoenfelder, K. Duffus *et al.*, 2016 Identifying Causal Genes at the Multiple Sclerosis Associated Region 6q23 Using Capture Hi-C. PLoS One 11**:** e0166923.
McDowall, A. W., J. M. Smith and J. Dubochet, 1986 Cryo-electron microscopy of vitrified chromosomes in situ. EMBO J 5**:** 1395-1402.
McGhee, J. D., W. I. Wood, M. Dolan, J. D. Engel and G. Felsenfeld, 1981 A 200 base pair region at the 5' end of the chicken adult beta-globin gene is accessible to nuclease digestion. Cell 27**:** 45-55.
McGovern, A., S. Schoenfelder, P. Martin, J. Massey, K. Duffus *et al.*, 2016 Capture Hi-C identifies a novel causal gene, IL20RA, in the pan-autoimmune genetic susceptibility region 6q23. Genome Biol 17**:** 212.
Medler, S., N. Al Husini, S. Raghunayakula, B. Mukundan, A. Aldea *et al.*, 2011 Evidence for a complex of transcription factor IIB with poly(A) polymerase and cleavage factor 1 subunits required for gene looping. J Biol Chem 286**:** 33709-33718.
Medler, S., and A. Ansari, 2015 Gene looping facilitates TFIIH kinase-mediated termination of transcription. Sci Rep 5**:** 12586.
Mews, P., B. M. Zee, S. Liu, G. Donahue, B. A. Garcia *et al.*, 2014 Histone methylation has dynamics distinct from those of histone acetylation in cell cycle reentry from quiescence. Mol Cell Biol 34**:** 3968-3980.
Miele, A., K. Bystricky and J. Dekker, 2009 Yeast silent mating type loci form heterochromatic clusters through silencer protein-dependent long-range interactions. PLoS Genet 5**:** e1000478.
Mifsud, B., F. Tavares-Cadete, A. N. Young, R. Sugar, S. Schoenfelder *et al.*, 2015 Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet 47**:** 598-606.
Mizuguchi, T., G. Fudenberg, S. Mehta, J. M. Belton, N. Taneja *et al.*, 2014 Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. Nature 516**:** 432-435.
Mukundan, B., and A. Ansari, 2011 Novel role for mediator complex subunit Srb5/Med18 in termination of transcription. J Biol Chem 286**:** 37053-37057.
Mukundan, B., and A. Ansari, 2013 Srb5/Med18-mediated termination of transcription is dependent on gene looping. J Biol Chem 288**:** 11384-11394.
Nagano, T., Y. Lubling, T. J. Stevens, S. Schoenfelder, E. Yaffe *et al.*, 2013 Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature 502**:** 59-64.
Nagano, T., Y. Lubling, E. Yaffe, S. W. Wingett, W. Dean *et al.*, 2015a Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. Nat Protoc 10**:** 1986-2003.
Nagano, T., C. Varnai, S. Schoenfelder, B. M. Javierre, S. W. Wingett *et al.*, 2015b Comparison of Hi-C results using in-solution versus in-nucleus ligation. Genome Biol 16**:** 175.

Nagy, P. L., M. L. Cleary, P. O. Brown and J. D. Lieb, 2003 Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin. Proc Natl Acad Sci U S A 100**:** 6364-6369.
Narendra, V., P. P. Rocha, D. An, R. Raviram, J. A. Skok *et al.*, 2015 CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science 347**:** 1017-1021.
Naumova, N., M. Imakaev, G. Fudenberg, Y. Zhan, B. R. Lajoie *et al.*, 2013 Organization of the mitotic chromosome. Science 342**:** 948-953.
Negre, N., C. D. Brown, P. K. Shah, P. Kheradpour, C. A. Morrison *et al.*, 2010 A comprehensive map of insulator elements for the Drosophila genome. PLoS Genet 6**:** e1000814.
Nishino, Y., M. Eltsov, Y. Joti, K. Ito, H. Takata *et al.*, 2012 Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. EMBO J 31**:** 1644-1653.
Nock, A., J. M. Ascano, M. J. Barrero and S. Malik, 2012 Mediator-regulated transcription through the +1 nucleosome. Mol Cell 48**:** 837-848.
Noll, M., and R. D. Kornberg, 1977 Action of micrococcal nuclease on chromatin and the location of histone H1. J Mol Biol 109**:** 393-404.
Noordermeer, D., M. Leleu, E. Splinter, J. Rougemont, W. De Laat *et al.*, 2011 The dynamic architecture of Hox gene clusters. Science 334**:** 222-225.
Nora, E. P., A. Goloborodko, A. L. Valton, J. Gibcus, A. Uebersohn *et al.*, 2016 Targeted degradation of CTCF decouples local insulation of chromosome domains from higher-order genomic compartmentalization. bioRxiv.
Nora, E. P., B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto *et al.*, 2012 Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485**:** 381-385.
O'Sullivan, J. M., S. M. Tan-Wong, A. Morillon, B. Lee, J. Coles *et al.*, 2004 Gene loops juxtapose promoters and terminators in yeast. Nat Genet 36**:** 1014-1018.
Olins, A. L., and D. E. Olins, 1974 Spheroid chromatin units (v bodies). Science 183**:** 330-332.
Perisic, O., R. Collepardo-Guevara and T. Schlick, 2010 Modeling studies of chromatin fiber structure as a function of DNA linker length. J Mol Biol 403**:** 777-802.
Petrenko, N., Y. Jin, K. H. Wong and K. Struhl, 2016 Mediator Undergoes a Compositional Change during Transcriptional Activation. Mol Cell 64**:** 443-454.
Phillips-Cremins, J. E., M. E. Sauria, A. Sanyal, T. I. Gerasimova, B. R. Lajoie *et al.*, 2013 Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell 153**:** 1281-1295.
Pickersgill, H., B. Kalverda, E. de Wit, W. Talhout, M. Fornerod *et al.*, 2006 Characterization of the Drosophila melanogaster genome at the nuclear lamina. Nat Genet 38**:** 1005-1014.
Plaschka, C., L. Lariviere, L. Wenzeck, M. Seizl, M. Hemann *et al.*, 2015 Architecture of the RNA polymerase II-Mediator core initiation complex. Nature 518**:** 376-380.
Pope, B. D., T. Ryba, V. Dileep, F. Yue, W. Wu *et al.*, 2014 Topologically associating domains are stable units of replication-timing regulation. Nature 515**:** 402-405.
Radman-Livaja, M., G. Ruben, A. Weiner, N. Friedman, R. Kamakaka *et al.*, 2011a Dynamics of Sir3 spreading in budding yeast: secondary recruitment sites and euchromatic localization. Embo J 30**:** 1012-1026.
Radman-Livaja, M., K. F. Verzijlbergen, A. Weiner, T. van Welsem, N. Friedman *et al.*, 2011b Patterns and mechanisms of ancestral histone protein inheritance in budding yeast. PLoS Biol 9**:** e1001075.
Raghuraman, M. K., E. A. Winzeler, D. Collingwood, S. Hunt, L. Wodicka *et al.*, 2001 Replication dynamics of the yeast genome. Science 294**:** 115-121.
Ramani, V., X. Deng, R. Qiu, K. L. Gunderson, F. J. Steemers *et al.*, 2017 Massively multiplex single-cell Hi-C. Nat Methods 14**:** 263-266.
Rando, O. J., 2007 Global patterns of histone modifications. Curr Opin Genet Dev 17**:** 94-99.
Rando, O. J., 2012 Combinatorial complexity in chromatin structure and function: revisiting the histone code. Curr Opin Genet Dev 22**:** 148-155.

Rando, O. J., and F. Winston, 2012 Chromatin and transcription in yeast. Genetics 190**:** 351-387.
Rao, S. S., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov *et al.*, 2014 A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159**:** 1665-1680.
Reddy, K. L., J. M. Zullo, E. Bertolino and H. Singh, 2008 Transcriptional repression mediated by repositioning of genes to the nuclear lamina. Nature 452**:** 243-247.
Ricci, M. A., C. Manzo, M. F. Garcia-Parajo, M. Lakadamyali and M. P. Cosma, 2015 Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. Cell 160**:** 1145-1158.
Risca, V. I., S. K. Denny, A. F. Straight and W. J. Greenleaf, 2017 Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping. Nature 541**:** 237-241.
Robinson, P. J., L. Fairall, V. A. Huynh and D. Rhodes, 2006 EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. Proc Natl Acad Sci U S A 103**:** 6506-6511.
Robinson, P. J., and D. Rhodes, 2006 Structure of the '30 nm' chromatin fibre: a key role for the linker histone. Curr Opin Struct Biol 16**:** 336-343.
Rouquette, J., C. Genoud, G. H. Vazquez-Nin, B. Kraus, T. Cremer *et al.*, 2009 Revealing the high-resolution three-dimensional network of chromatin and interchromatin space: a novel electron-microscopic approach to reconstructing nuclear architecture. Chromosome Res 17**:** 801-810.
Routh, A., S. Sandin and D. Rhodes, 2008 Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. Proc Natl Acad Sci U S A 105**:** 8872-8877.
Rufiange, A., P. E. Jacques, W. Bhat, F. Robert and A. Nourani, 2007 Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. Mol Cell 27**:** 393-405.
Rutledge, M. T., M. Russo, J. M. Belton, J. Dekker and J. R. Broach, 2015 The yeast genome undergoes significant topological reorganization in quiescence. Nucleic Acids Res 43**:** 8299-8313.
Rydberg, B., W. R. Holley, I. S. Mian and A. Chatterjee, 1998 Chromatin conformation in living cells: support for a zig-zag model of the 30 nm chromatin fiber. J Mol Biol 284**:** 71-84.
Sadeh, R., R. Launer-Wachs, H. Wandel, A. Rahat and N. Friedman, 2016 Elucidating Combinatorial Chromatin States at Single-Nucleosome Resolution. Mol Cell 63**:** 1080-1088.
Sanborn, A. L., S. S. Rao, S. C. Huang, N. C. Durand, M. H. Huntley *et al.*, 2015 Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A 112**:** E6456-6465.
Sandhu, K. S., G. Li, H. M. Poh, Y. L. Quek, Y. Y. Sia *et al.*, 2012 Large-scale functional organization of long-range chromatin interaction networks. Cell Rep 2**:** 1207-1219.
Sanyal, A., B. R. Lajoie, G. Jain and J. Dekker, 2012 The long-range interaction landscape of gene promoters. Nature 489**:** 109-113.
Schalch, T., S. Duda, D. F. Sargent and T. J. Richmond, 2005 X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature 436**:** 138-141.
Scheffer, M. P., M. Eltsov, J. Bednar and A. S. Frangakis, 2012 Nucleosomes stacked with aligned dyad axes are found in native compact chromatin in vitro. J Struct Biol 178**:** 207-214.
Scheffer, M. P., M. Eltsov and A. S. Frangakis, 2011 Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei. Proc Natl Acad Sci U S A 108**:** 16992-16997.
Schermelleh, L., I. Solovei, D. Zink and T. Cremer, 2001 Two-color fluorescence labeling of early and mid-to-late replicating chromatin in living cells. Chromosome Res 9**:** 77-80.
Scherrer, R., L. Louden and P. Gerhardt, 1974 Porosity of the yeast cell wall and membrane. J Bacteriol 118**:** 534-540.

Schlick, T., J. Hayes and S. Grigoryev, 2012 Toward convergence of experimental studies and theoretical modeling of the chromatin fiber. J Biol Chem 287**:** 5183-5191.
Schmitt, A. D., M. Hu and B. Ren, 2016 Genome-wide mapping and analysis of chromosome architecture. Nat Rev Mol Cell Biol 17**:** 743-755.
Schoenfelder, S., M. Furlan-Magaril, B. Mifsud, F. Tavares-Cadete, R. Sugar *et al.*, 2015 The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res 25**:** 582-597.
Schwarzer, W., N. Abdennur, A. Goloborodko, A. Pekowska, G. Fudenberg *et al.*, 2016 Two independent modes of chromosome organization are revealed by cohesin removal. bioRxiv.
Seitan, V. C., A. J. Faure, Y. Zhan, R. P. McCord, B. R. Lajoie *et al.*, 2013 Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Res 23**:** 2066-2077.
Sekinger, E. A., Z. Moqtaderi and K. Struhl, 2005 Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. Mol Cell 18**:** 735-748.
Sexton, T., E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc *et al.*, 2012 Three-dimensional folding and functional organization principles of the Drosophila genome. Cell 148**:** 458-472.
Shogren-Knaak, M., H. Ishii, J. M. Sun, M. J. Pazin, J. R. Davie *et al.*, 2006 Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science 311**:** 844-847.
Shopland, L. S., C. R. Lynch, K. A. Peterson, K. Thornton, N. Kepper *et al.*, 2006 Folding and organization of a contiguous chromosome region according to the gene distribution pattern in primary genomic sequence. J Cell Biol 174**:** 27-38.
Simonis, M., P. Klous, E. Splinter, Y. Moshkin, R. Willemsen *et al.*, 2006 Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38**:** 1348-1354.
Singh, B. N., and M. Hampsey, 2007 A transcription-independent role for TFIIB in gene looping. Mol Cell 27**:** 806-816.
Sinha, I., M. Wiren and K. Ekwall, 2006 Genome-wide patterns of histone modifications in fission yeast. Chromosome Res 14**:** 95-105.
Sofueva, S., E. Yaffe, W. C. Chan, D. Georgopoulou, M. Vietri Rudan *et al.*, 2013 Cohesin-mediated interactions organize chromosomal domain architecture. EMBO J 32**:** 3119-3129.
Solovei, I., M. Kreysing, C. Lanctot, S. Kosem, L. Peichl *et al.*, 2009 Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. Cell 137**:** 356-368.
Song, F., P. Chen, D. Sun, M. Wang, L. Dong *et al.*, 2014 Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. Science 344**:** 376-380.
Spain, M. M., S. A. Ansari, R. Pathak, M. J. Palumbo, R. H. Morse *et al.*, 2014 The RSC complex localizes to coding sequences to regulate Pol II and histone occupancy. Mol Cell 56**:** 653-666.
Suganuma, T., and J. L. Workman, 2008 Crosstalk among Histone Modifications. Cell 135**:** 604-607.
Symmons, O., V. V. Uslu, T. Tsujimura, S. Ruf, S. Nassari *et al.*, 2014 Functional and topological characteristics of mammalian regulatory domains. Genome Res 24**:** 390-400.
Szilard, R. K., P. E. Jacques, L. Laramee, B. Cheng, S. Galicia *et al.*, 2010 Systematic identification of fragile sites via genome-wide location analysis of gamma-H2AX. Nat Struct Mol Biol 17**:** 299-305.
Taddei, A., and S. M. Gasser, 2012 Structure and function in the budding yeast nucleus. Genetics 192**:** 107-129.
Takebayashi, S., V. Dileep, T. Ryba, J. H. Dennis and D. M. Gilbert, 2012 Chromatin-interaction compartment switch at developmentally regulated chromosomal domains reveals an unusual principle of chromatin folding. Proc Natl Acad Sci U S A 109**:** 12574-12579.

Tan-Wong, S. M., H. D. Wijayatilake and N. J. Proudfoot, 2009 Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. Genes Dev 23**:** 2610-2624.

Tan-Wong, S. M., J. B. Zaugg, J. Camblong, Z. Xu, D. W. Zhang *et al.*, 2012 Gene loops enhance transcriptional directionality. Science 338**:** 671-675.

Tan, M., H. Luo, S. Lee, F. Jin, J. S. Yang *et al.*, 2011 Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell 146**:** 1016-1028.

Therizols, P., R. S. Illingworth, C. Courilleau, S. Boyle, A. J. Wood *et al.*, 2014 Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. Science 346**:** 1238-1242.

Thoma, F., T. Koller and A. Klug, 1979 Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J Cell Biol 83**:** 403-427.

Tremethick, D. J., 2007 Higher-order structures of chromatin: the elusive 30 nm fiber. Cell 128**:** 651-654.

Tsai, K. L., C. Tomomori-Sato, S. Sato, R. C. Conaway, J. W. Conaway *et al.*, 2014 Subunit architecture and functional modular rearrangements of the transcriptional mediator complex. Cell 157**:** 1430-1444.

Tsankov, A. M., D. A. Thompson, A. Socha, A. Regev and O. J. Rando, 2010 The role of nucleosome positioning in the evolution of gene regulation. PLoS Biol 8**:** e1000414.

Tsubota, T., C. E. Berndsen, J. A. Erkmann, C. L. Smith, L. Yang *et al.*, 2007 Histone H3-K56 Acetylation Is Catalyzed by Histone Chaperone-Dependent Complexes. Mol Cell 25**:** 703-712.

Turner, B. M., 1993 Decoding the nucleosome. Cell 75**:** 5-8.

Ulianov, S. V., E. E. Khrameeva, A. A. Gavrilov, I. M. Flyamer, P. Kos *et al.*, 2016 Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. Genome Res 26**:** 70-84.

Uuskula-Reimand, L., H. Hou, P. Samavarchi-Tehrani, M. V. Rudan, M. Liang *et al.*, 2016 Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. Genome Biol 17**:** 182.

van Bakel, H., K. Tsui, M. Gebbia, S. Mnaimneh, T. R. Hughes *et al.*, 2013 A compendium of nucleosome and transcript profiles reveals determinants of chromatin architecture and transcription. PLoS Genet 9**:** e1003479.

van de Peppel, J., N. Kettelarij, H. van Bakel, T. T. Kockelkorn, D. van Leenen *et al.*, 2005 Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets. Mol Cell 19**:** 511-522.

Vietri Rudan, M., C. Barrington, S. Henderson, C. Ernst, D. T. Odom *et al.*, 2015 Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep 10**:** 1297-1309.

Vogelmann, J., A. Le Gall, S. Dejardin, F. Allemand, A. Gamot *et al.*, 2014 Chromatin insulator factors involved in long-range DNA interactions and their role in the folding of the Drosophila genome. PLoS Genet 10**:** e1004544.

Voong, L. N., L. Xi, A. C. Sebeson, B. Xiong, J. P. Wang *et al.*, 2016 Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. Cell 167**:** 1555-1570 e1515.

Wang, C., C. Liu, D. Roqueiro, D. Grimm, R. Schwab *et al.*, 2015 Genome-wide analysis of local chromatin packing in Arabidopsis thaliana. Genome Res 25**:** 246-256.

Wang, S., J. H. Su, B. J. Beliveau, B. Bintu, J. R. Moffitt *et al.*, 2016 Spatial organization of chromatin domains and compartments in single chromosomes. Science 353**:** 598-602.

Watanabe, S., M. Radman-Livaja, O. J. Rando and C. L. Peterson, 2013 A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme. Science 340**:** 195-199.

Weiner, A., H. V. Chen, C. L. Liu, A. Rahat, A. Klien *et al.*, 2012 Systematic dissection of roles for chromatin regulators in a yeast stress response. PLoS Biol 10**:** e1001369.
Weiner, A., T. H. Hsieh, A. Appleboim, H. V. Chen, A. Rahat *et al.*, 2015 High-resolution chromatin dynamics during a yeast stress response. Mol Cell 58**:** 371-386.
Weiner, A., A. Hughes, M. Yassour, O. J. Rando and N. Friedman, 2010 High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Res 20**:** 90-100.
Weiner, A., D. Lara-Astiaso, V. Krupalnik, O. Gafni, E. David *et al.*, 2016 Co-ChIP enables genome-wide mapping of histone mark co-occurrence at single-molecule resolution. Nat Biotechnol 34**:** 953-961.
Wijchers, P. J., P. H. Krijger, G. Geeven, Y. Zhu, A. Denker *et al.*, 2016 Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments. Mol Cell 61**:** 461-473.
Williams, S. P., B. D. Athey, L. J. Muglia, R. S. Schappe, A. H. Gough *et al.*, 1986 Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length. Biophys J 49**:** 233-248.
Williamson, I., S. Berlivet, R. Eskeland, S. Boyle, R. S. Illingworth *et al.*, 2014 Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. Genes Dev 28**:** 2778-2791.
Wong, H., J. M. Victor and J. Mozziconacci, 2007 An all-atom model of the chromatin fiber containing linker histones reveals a versatile structure tuned by the nucleosomal repeat length. PLoS One 2**:** e877.
Woodcock, C. L., L. L. Frado and J. B. Rattner, 1984 The higher-order structure of chromatin: evidence for a helical ribbon arrangement. J Cell Biol 99**:** 42-52.
Woodcock, C. L., S. A. Grigoryev, R. A. Horowitz and N. Whitaker, 1993 A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. Proc Natl Acad Sci U S A 90**:** 9021-9025.
Worcel, A., S. Strogatz and D. Riley, 1981 Structure of chromatin and the linking number of DNA. Proc Natl Acad Sci U S A 78**:** 1461-1465.
Xi, Y., J. Yao, R. Chen, W. Li and X. He, 2011 Nucleosome fragility reveals novel functional states of chromatin and poises genes for activation. Genome Res 21**:** 718-724.
Xu, Y. Z., C. Kanagaratham and D. Radzioch, 2013 Chromatin Remodelling During Host-Bacterial Pathogen Interaction, pp. Ch. 08 in *Chromatin Remodelling*, edited by D. Radzioch. InTech, Rijeka.
Xu, Z., W. Wei, J. Gagneur, F. Perocchi, S. Clauder-Münster *et al.*, 2009 Bidirectional promoters generate pervasive transcription in yeast. Nature 457**:** 1033-1037.
Yadon, A. N., B. N. Singh, M. Hampsey and T. Tsukiyama, 2013 DNA looping facilitates targeting of a chromatin remodeling enzyme. Mol Cell 50**:** 93-103.
Yuan, G. C., Y. J. Liu, M. F. Dion, M. D. Slack, L. F. Wu *et al.*, 2005 Genome-scale identification of nucleosome positions in S. cerevisiae. Science 309**:** 626-630.
Zentner, G. E., and S. Henikoff, 2014 High-resolution digital profiling of the epigenome. Nat Rev Genet 15**:** 814-827.
Zentner, G. E., S. Kasinathan, B. Xin, R. Rohs and S. Henikoff, 2015 ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. Nat Commun 6**:** 8733.
Zhang, Y., C. H. Wong, R. Y. Birnbaum, G. Li, R. Favaro *et al.*, 2013 Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature 504**:** 306-310.
Zhang, Z., and B. F. Pugh, 2011 High-resolution genome-wide mapping of the primary structure of chromatin. Cell 144**:** 175-186.
Zhang, Z., C. J. Wippo, M. Wal, E. Ward, P. Korber *et al.*, 2011 A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. Science 332**:** 977-980.

Zhao, Z., G. Tavoosidana, M. Sjolinder, A. Gondor, P. Mariano *et al.*, 2006 Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet 38**:** 1341-1347.

Zhu, X., B. Liu, J. O. Carlsten, J. Beve, T. Nystrom *et al.*, 2011 Mediator influences telomeric silencing and cellular life span. Mol Cell Biol 31**:** 2413-2421.

Zimmer, C., and E. Fabre, 2011 Principles of chromosomal organization: lessons from yeast. J Cell Biol 192**:** 723-733.

Zuin, J., J. R. Dixon, M. I. van der Reijden, Z. Ye, P. Kolovos *et al.*, 2014 Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc Natl Acad Sci U S A 111**:** 996-1001.