2014-12-09

The Recombination Enhancer Modulates the Conformation of Chr. III in Budding Yeast: A Dissertation

Jon-Matthew Belton
University of Massachusetts Medical School

THE RECOMBINATION ENHANCER MODULATES THE CONFORMATION OF CHR. III IN BUDDING YEAST

A Dissertation Presented

By

Jon-Matthew Belton

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

DECEMBER 9th 2014

BIOMEDICAL SCIENCES

THE RECOMBINATION ENHANCER MODULATES THE CONFORMATION OF CHR. III IN BUDDING YEAST

Dissertation Presented By

Jon-Matthew Belton

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

_____
Job Dekker, Ph.D., Thesis Advisor

_____
James Broach, Ph.D., Member of Committee

_____
Anthony Imbalzano, Ph.D., Member of Committee

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

_____
Marian Walhout, 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

_____
Oliver Rando, M.D., 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

December 9, 2014

DEDICATION

I would like to thank all of the teachers, coaches, and family members whom cared enough to push me to pursue my dreams. This thesis is dedicated to you.

ABSTRACT

A hierarchy of different chromosome conformations plays a role in many biological systems. These conformations contribute to the regulation of gene expression, cellular development, chromosome transmission, and defects can lead to human disease. The highest functional level of this hierarchy is the partitioning of the genome into compartments of active and inactive chromatin domains (1's -10's Mb). These compartments are further partitioned into Topologically Associating Domains (TADs) that spatially cluster co-regulated genes (100's kb – 1's Mb). The final level that has been observed is long range loops formed between regulatory elements and promoters (10's kb – 100's Mb). At all of these levels, mechanisms that establish these conformations remain poorly understood. To gain new insights into processes that determine chromosome folding I used the mating type switching system in budding yeast to study the chromosome conformation at length scales analogous to looping interaction. I specifically examined the role in chromosome conformation in the mating type switching system. Budding yeast cells can have two sexes: *MAT***a** and *MAT*α. The mating types are determined by allele-specific expression of the *MAT* locus on chromosome III. The *MAT***a** allele encodes for transcription factors responsible for the *MAT***a** mating type and the *MAT*α allele encodes transcription factors responsible for the *MAT*α mating type. Yeast cells can switch their mating type by a process that repairs a break at *MAT* using one of two silent loci, *HML* or *HMR*, as a donor to convert the allele at the *MAT* locus. When *MAT***a**

cells switch they prefer to use *HML*, which contains the *MAT*α allele, located at the end of the left arm. *MAT*α cells prefer to use *HMR*, which contains the *MAT***a** allele, located on the end of the right arm of chromosome III. The sequences of the *HM* loci are not important for donor preference. Instead the cell chooses the donor on the left arm in *MAT***a** cells and chooses the donor on the right arm in *MAT*α cells. This lack of sequence specificity has led to the hypothesis that the conformation of the chromosome may play a role in donor preference. I found that the conformation of chromosome III is, indeed, different between the two mating types. In *MAT***a** cells the chromosomes displays a more crumpled conformation in which the left arm of the chromosome interacts with a large region of the right arm which includes the centromere and the *MAT* locus. In *MAT*α cells, on the other hand, the left arm of the chromosomes displays a more extend conformation. I found that the Recombination Enhancer (RE), which enhances recombination along the left arm of the chromosome in *MAT***a** cells, is responsible for these mating type-specific conformations. Deleting the RE affects the conformation of the chromosomes in both *MAT***a** and *MAT*α cells. The left portion of the RE, which is essential for donor preference during the switching reaction in *MAT***a** cells, does not contribute to the conformation in *MAT***a**. This region does have a minor effect on the conformation in *MAT*α cells. However, I found that the right portion of the RE is responsible for the conformation of chromosome III in both mating types prior to initiation of switching. This work

demonstrates that chromosome conformation is determined by specific cis regulatory elements that drive cell-type specific chromosome conformation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABREVIATIONS

2D	two-dimensional
3C	Chromosome Conformation Capture
3D	three-dimensional
4C	3C-on-Chip or Circular 3C
5C	Chromosome Conformation Capture Carbon Copy
ATP	Adenosine 5’ Triphosphate
bp	base pair
BSA	Bovine Serum Albumin
cen	centromere
CFTR	Cystic Fibrosis Transmembrane Receptor
ChIP	Chromatin Immuno-Precipitation
chr	chromosome
CID	Chromatin Interaction Domain
Cl	chlorine
dATP	2’ Deoxyribose Adenosine 5’ Triphosphate
dGTP	2’ Deoxyribose Guanosine 5’ Triphosphate

dTTP	2’ Deoxyribose Thymidine 5’ Triphosphate
dCTP	2’ Deoxyribose Cytosine 5’ Triphosphate
DNA	2’ Deoxyribonucleic Acid
dNTP	2’ Deoxyribose Nucleotide 5’ Triphosphate
DPS1	Donor Preference Site 1, located in the left part of the RE
DPS2	Donor Preference Site 2, located in the right part of the RE
EB	Elution Buffer
EDTA	Ethylenediaminetetraacetic acid
FISH	Fluorescent In Situ Hybridization
*fkh1*Δ	Deletion of the Fkh1 gene by replacement with KanMX
Hi-C	High throughput chromosome conformation assay
HML	Homothallic Left
HMR	Homothallic Right
HO	The endonuclease that cleaves the MAT locus during switching.
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End Joining
MAT	The mating type locus on chromosome III.

*MAT***a**	The A allele at the MAT locus and the A sex of yeast cells
MAT**a**1p	One of the two protein expressed from the *MAT***a** locus
MAT**a**2p	One of the two protein expressed from the *MAT***a** locus
*MAT*α	The α allele at the MAT locus and the α sex of yeast cells
MATα1p	One of the two protein expressed from the *MAT*α locus
MATα2p	One of the two protein expressed from the *MAT*α locus
IHI	Interactive Heterochromatin Islands
LAD	Lamin Associated Domain
LCR	Locus Control Region of the beta-globin locus
Mcm1	Mini-Chromosome Maintenance
RE	Recombination Enhancer
reΔ-left	Deletion of the left portion of the RE
reΔ-right	Deletion of the right portion of the RE
reΔ-whole	Deletion of the entire portion of the RE
SBC	Swi4-Swi6 Binding Cite
SBF	SBC Binding Factor
SIR	Silent Information Regulator

SMC	Structural Maintenance of Chromosomes
SPB	Spindle Pole Body
SPRI	Solid Phase Reversible Immobilization
TAD	Topologically Associated Domain

List of Publications

CHAPTER II Belton, J.-M., R. P. McCord, et al. (2012). "Hi–C: A comprehensive technique to capture the conformation of genomes." Methods **58**(3): 268-276.

CHAPTER III Belton, J.-M., Dekker, J. “Measuring Chromatin Structure in Budding Yeast.” Budding Yeast: A Laboratory Manual. Cold Spring Harbor Laboratory Press. – *Accepted for Publication*

CHAPTER I

Introduction

Contributions

The text for the section " Yeast Nuclear Organization" are from: Belton, J.-M., Dekker, J. "Measuring Chromatin Structure in Budding Yeast." Budding Yeast: A Laboratory Manual. Cold Spring Harbor Laboratory Press. – *Accepted for Publication.* The authors are listed below: Jon-Matthew Belton, Job Dekker

Why is studying chromosome conformation important

Chromatin is a very large and complex set of molecules that resides in a very confined volume of the nucleus. The distribution of chromatin in the nucleus is not random (sanyalMonneron and Bernhard 1969; Lieberman-Aiden et al. 2009a; Dixon et al. 2012a; Nora et al. 2012). A hierarchy of structures has been observed with various technologies in various organisms. This hierarchy has been shown to be involved in the regulation of genes through long range interactions between promoters and enhancer elements (Sanyal et al. 2012). Furthermore, genes that are co-regulated are often in local domains that share regulator elements (Andrey et al. 2013; Le Dily et al. 2014). On a higher level, domains that are active associate with one another while domains that are inactive associate with one another. This partitions the genome, spatially, into compartments of active and inactive chromatin (Lieberman-Aiden et al. 2009a). Defects in chromatin structure have been associated with Hutchinson-Guilford Progeria syndrome (McCord et al. 2013) and the formation of translocations, since translocations form most often with regions that are in the same chromatin compartment (Zhang et al. 2012a). The mechanisms behind the establishment of these structures is not fully understood. Understanding the structural organization of chromosomes is essential to fully grasp the complexity of gene regulation and the etiology of pathologies of the cell nucleus.

Mating type switching in budding yeast as a model system

The Mating type switching system in budding yeast has been used to lay the groundwork for the understanding of many molecular phenomena. The fact that budding yeast is a genetically tractable organism and that the mating type switching process provides easily screenable phenotypes makes this system ideal for experimentation. Experiments on the mating type switching process have led to the discovery of mechanisms of gene silencing, chromatin remodeling, cell type specific expression, cell cycle regulation, and DNA break repair. In this thesis, the differences between the two mating types is exploited to study chromosome conformation.

Budding yeast has a genome size of ~12Mb distributed over 16 chromosomes, is genetically tractable and its genome can easily be manipulated. This, coupled with the vast depth of knowledge know about the physiology of budding yeast, allows for the analysis of chromatin interactions in contexts that are not readily available in higher eukaryotes, including an array of unique environmental conditions. Further, the roles of specific proteins suspected to be involved in chromatin organization can be straightforwardly studied. In addition, yeast cultures are easily synchronized to study changes in chromosome conformation during the cell cycle.

Mating type switching

Budding yeast can switch their mating types. Presumably they do this in order to keep the haploid population equally distributed between the two mating types. Mating type switching involves a gene conversion event at the *MAT* locus on Chr. III that switches the allele and thus changes the expression state of the cell (**Fig. 1.1**). Copies of the *MAT* alleles are held in two silenced/hidden loci. Homothallic Mating Left (*HML*) contains the **α** allele and the Homothallic Mating Right (*HMR*) contains the **a** allele in wild type strains. The switching process occurs in G1 mother cells that have previously divided and is initiated by cleavage of the *MAT* locus by HO, a site specific endonuclease. The break at the *MAT* locus is repaired through the mechanism of Synthesis Dependent Strand Annealing (SDSA) that uses the sequence in one of the *HM* loci as a template to repair the break at *MAT*. This process copies the allele from the donor *HM* locus that is being used as a template into the *MAT* locus.

In wild type *MAT***a** cells, the break at *MAT* is repaired using the donor sequence at *HML* as a template, whereas *MAT***α** cells repair the break using the donor sequence at *HMR* (Weiler and Broach 1992a; Wu et al. 1998a; Sun et al. 2002b; Ercan and Simpson 2004; Ercan et al. 2005). This process occurs with 80-90% efficiency in both cases (Weiler and Broach 1992a; Li et al. 2012). Interestingly, the choice of *HML* or *HMR* is not affected by the specific allele at the two loci (Coïc et al. 2006; Li et al. 2012). For instance, if the two alleles are switched such that *HML* contains the **a**-allele and *HMR* contains the α-allele, *MAT***a** cells will still choose *HML* as the donor and *MAT*α cells with still choose

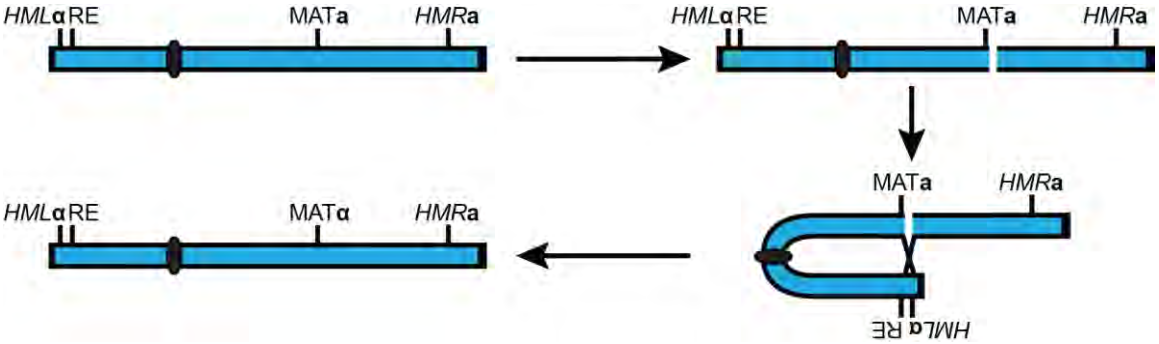


Figure 1.1. Schematic of mating type switching. Switching is induced by a double strand break at *MAT* created by HO. One donor is chosen to repair the break via a gene conversion event.

HMR as the donor (Weiler and Broach 1992a). One hypothesis is that sequences adjacent to the *MAT* allele at the *HM* loci must confer some sort of cell type specificity. However, the silencer sequences that comprise the *HM* loci are identical between *HML* and *HMR*, and when *HML* is replaced by *HMR*, *MAT***a** cells choose the donor on the left arm of chromosome while *MAT*α cells choose the donor on the right arm of chromosome III. These observations indicate that donor preference is not directed by sequences at either *HML* or *HMR*.

Cell type specific expression in yeast

The two haploid cell types in budding yeast termed *MAT***a** and *MAT*α are delineated by cell type specific expression of the *MAT* locus (Klar 1987). In *MAT***a** cells the "A" allele is expressed whereas in *MAT*α cells the "alpha" allele is expressed (Tatchell et al. 1981). The "A" allele produces two transcripts called MAT**a**1 and MAT**a**2 (Tatchell et al. 1981). Apparently, neither of the *MAT***a** proteins are necessary for expression of the "A" expression state. In fact, this state appears to be the default expression state, with the constitutive transcription factor Mcm1p responsible for the expression of A-specific genes (Haber 2012a). It is interesting that although there is no known activity of the *MAT***a** proteins in *MAT***a** cells, they are nonetheless still produced. This could mean that there is yet to be a function found for these proteins in *MAT***a** cells or that the expression of the *MAT***a** locus keeps the locus open and accessible for

cleavage by HO. MAT**a**1p does form a heterodimer with *MAT*α2p in diploid cells which express both of the *MAT* alleles. This complex represses the expression of haploid specific genes.

The *MAT*α expression state is different from the default state and is established by *MAT*α1p and *MAT*α2p. *MAT*α1p is responsible for activating the expression of *MAT*α specific genes. On the contrary, *MAT*α2p forms a heterodimer with Mcm1p which represses *MAT***a** specific genes (Hagen et al. 1993).

Genome-wide analysis shows that the two haploid and the diploid expression states differ from one another in only a few genes. Of the ~6,000 genes in the yeast genome, there are a total of 6 *MAT***a**-specific genes, 5 *MAT*α specific genes, and 19 haploid specific genes (Galgoczy et al. 2004). Haploids differ from diploids in their ability to mate with other haploids, their ability to switch their mating type, axial budding, their inability to undergo meiosis and sporulation, and their inability to use Non-homologous End Joining (NHEJ) as a repair pathway (Haber 2012a). *MAT***a** and *MAT*α cells differ from one another in which sex they choose to mate with and in which donor they use to repair the break at *MAT* during mating type switching. It is striking that so few genes can lead to these biological differences between the three cell types.

Components of the mating type switching system

The *MAT* locus is composed of 5 regions, known as; “W”, “X”, “Y”, “Z1”, “Z2” (Haber 1992). The Y region contains the coding sequence for the *MAT* genes and is denoted either Y**a** or Yα. Regions X and Z1 flank the Y region and are homologous to regions in *HML* and *HMR* (Haber 1998b). Region W is upstream of region X and region Z2 is down stream of Z1 (Haber 1998b). These two regions provide additional homology between *MAT* and *HML*.

HML and *HMR* both contain silent copies of the *MAT* locus. These regions are silenced to prevent expression of the *MAT* alleles they contain and to keep these regions refractory to cleavage by HO. *HML* and *HMR* are each flanked by silencers, called *HML*-E, *HML*-I and *HMR*-E and *HMR*-I respectively (Weiss and Simpson 1998; Ravindra et al. 1999). The silencers create an array of ordered nucleosomes along the length of the hidden locus they flank, which is different than the nucleosome positioning at the *MAT* locus (Weiss and Simpson 1998; Ravindra et al. 1999; Ercan and Simpson 2004). The nucleosome positioning at both *HML* and *HMR* is dependent on Sir3 and to a lesser extent on Sir1 (Weiss and Simpson 1998; Ravindra et al. 1999). On a historical note, genetic screens for mutants that lost silencing of the *HM* loci is how the SIR (Silent Information Regulators) genes were first identified (Rine and Herskowitz 1987). Complete silencing of *HML* and *HMR* requires Sir1, Sir2, Sir3, and Sir4 (Haber and George 1979; Miller and Nasmyth 1984; Rine and Herskowitz 1987). The E silencer is essential for silencing of *HML* and *HMR* and contains an ACS (ARS Consensus Region), an ABF1 biding site and a RAP1 binding site

(Kimmerly et al. 1988; Fox et al. 1997). Sir1, Sir3, and Sir4 bind to RAP1 and ABF1 in this region (Kimmerly et al. 1988). The I silencer is important but not essential and contains an ACS and an ABF1 binding site (Buchman et al. 1988). RAD54 is necessary for remodeling the nucleosomes in *HML* during strand invasion of the switching process (Hicks et al. 2011). Fkh1 and Fkh2 proteins also bind *HML* (Ostrow et al. 2014). Binding of these proteins is likely mediated by a direct interaction with ORC (Knott et al. 2012).

The Recombination Enhancer enhances recombination, in general, along the left arm of Chr. III in *MAT***a** cells and is necessary for usage of the left arm donor to repair the break at *MAT* in *MAT***a** cells (Wu and Haber 1996b; Szeto et al. 1997a). The RE is characterized by having an ordered nucleosome array in *MAT*α cells whereas it has unphased nucleosomes in *MAT***a** cells (Weiss and Simpson 1997; Ercan and Simpson 2004). In *MAT***a** cells the RE is inactivated by the repressor complex Mcm1p/*MATα2p* as well as Tup1/Ssn6 (Szeto and Broach 1997; Weiss and Simpson 1997). Mcm1p/*MATα2p* binds to two sites in the RE (Weiss and Simpson 1997). Several non-coding transcripts are produced from the RE (Ercan et al. 2005).

The RE was first discovered by deleting portions of the left arm of Chr. III in order to screen for cis regulatory elements enhancing usage of the left arm donor during switching (Wu and Haber 1996b). A 2.5 kb deletion reduced the preference for usage of the left arm donor in *MAT***a** cells to background levels (Wu and Haber 1996b). This 2.5 kb region was further probed with smaller but

still large (800 – 2,300 bp) deletions. A deletion that included the leftmost Mcm1p/*MATα2p* site reduced left arm usage to background levels, but, deletions that included the right Mcm1p/*MATα2p* site had only a minor effect on left arm usage. Deletion of regions downstream of the leftmost Mcm1p/*MATα2p* site also significantly reduced left arm usage to just above background levels. This indicates that both the leftmost Mcm1p/*MATα2p* binding site and the region just downstream of this site are essential for RE function (Wu and Haber 1996b). This region is known as the 700bp minimal RE. These experiments were done in *MAT***a** cells and were designed to ascertain which parts of the RE are required to "activate" the RE. The effect of these deletions in *MAT*α cells is unknown. Since Mcm1p/*MATα2p* is a repressor complex, it is possible that these sites are also necessary to "shut down" the RE in *MAT*α cells.

The 700bp minimal RE contains several conserved regions called region "A", "B", "C", "D", and "E" (Wu et al. 1998b; Sun et al. 2002b; Coïc et al. 2006). Regions A, B, D, and E contain Fkh1 binding sites that bind Fkh1 in *MAT***a** cells with regions D and E containing arrays of several Fkh1 binding sites in a row (Wu et al. 1998b; Sun et al. 2002b; Coïc et al. 2006). Region C contains the Mcm1p/*MATα2p* operator as well as an SCB (Swi4-Swi6 cell cycle box) site which is bound by SBF (Swi4-Swi6 cell cycle box binding factor) (Wu et al. 1998b; Sun et al. 2002b; Coïc et al. 2006). Many combinations of these regions have been tested for the frequency by which the left arm donor is used in *MAT***a** cells (Sun et al. 2002b). The studies showed that first, region B is not essential

for switching (Sun et al. 2002b). Second, any combination of regions that results in long arrays of Fkh1 binding sites usually uses the left arm just as much as the wild type RE configuration (Sun et al. 2002b). Experiments in which the RE has been replaced with LexA biding sites that bind either LexA-Fkh1 or LexA-FHA domain from Fkh1 show a partial rescue of left arm usage compared to the RE deletion alone (Li et al. 2012). Third, region C, the Mcm1p/*MATα2p* and SBF operator, is not essential for switching (Sun et al. 2002b). These results suggest that the mechanism behind the physical act of switching itself is centered around the activity of Fkh1 and that the operators in region C are just used as a molecular switch to control cell type and cell cycle specific RE activity. Presumably, without the Mcm1 binding site the arrays of Fkh1p sites would be made inaccessible by phased nucleosomes in both mating types.

Cell cycle regulation of mating type switching

The Mating type switching system is tightly regulated by the cell cycle. Any G1 haploid cell that has previously budded, can switch its mating type, but new daughter cells cannot (Strathern and Herskowitz 1979). This keeps the population equally distributed with *MAT***a** and *MAT*α cells and allows for the maximum probability of mating to reconstitute the diploid state. Mating type switching is initiated by endonucleolytic cleavage of the *MAT* locus by the endonuclease HO. HO is expressed at the end of G1 phase, right after the

commitment to S-phase. Compared to most yeast genes, HO has a complex and modular promoter. The promoter is roughly 1.4 kb in length and contains two regions: URS1 (Upstream Regulatory Sequence 1) and URS2 (Upstream Regulatory Sequence 2) (Nasmyth 1987; Nasmyth 1993). URS1 is essential for mother cell specificity (Nasmyth 1987). Swi5 binds to this portion of the promoter in mother cells and helps activate transcription of HO (Nasmyth 1987). Ash1p is transported into the daughter cell and represses Swi5 which represses HO expression in these cells (Bobola et al. 1996). The second region, URS2, is essential for cell cycle control of HO. URS2 is bound by SBF (Swi4-Swi6) which is activated by Cdk1 (Nasmyth 1993; Haber 2012a). The promoter of HO also has Mcm1p/*MATα2p* binding sites which serve to repress the expression of HO in diploid cells. To prevent HO from cleaving the *MAT* locus more than once, HO is rapidly exported from the nucleus and degraded by the proteasome (Kaplun et al. 2003). This fine balancing of expression and degradation, the fact that there is only one copy of *MAT* that is cleavable, and the fact that it takes about an hour to repair a break at *MAT*, limits HO to about one cut per cell cycle.

HO expression is not the only part of the mating type switching system that is controlled by the cell cycle. Use of the left arm as a donor to repair the break at *MAT* appears to be a cell cycle coupled process. The RE is controlled by two independent cell cycle linked mechanisms. First, SBF (Swi4-Swi6) binds to the left part of the RE, just upstream of DPS1, during the G1/S transition and is necessary for full RE activity (Coïc et al. 2006). Additionally, Fkh1 and Fkh2

binds to several arrays of binding sites in the RE (Sun et al. 2002b; Coïc et al. 2006) Recent, high throughput microarray data of the left arm of Chr. III, has shown that both Fkh1 and Fkh2 bind to the RE throughout the cell cycle (Ostrow et al. 2014). This may keep the RE active throughout the cell cycle. It is unclear why Fkh2 binds to the RE since it is not necessary for donor preference (Sun et al. 2002b). In fact, Fkh2 seems to have a minor antagonistic effect on donor preference since deleting Fkh2, in both wild type and deltafkh1 strains, slightly increases left arm usage (Sun et al. 2002b). This could be due to competition for binding sites since the consensus sequence of Fkh1 and Fkh2 are similar. Mcm1p does not seem to be essential for the cell cycle regulation of the RE since it binds to the RE constitutively throughout the cell cycle (Ercan et al. 2005; Coïc et al. 2006).

Arrest of cells at either the G1/S or G2/M transitions cause a decrease in left arm usage (Coïc et al. 2006). This is further evidence that there are two independent cell cycle linked mechanisms that activate the RE in *MAT***a** cells. Interestingly, although arrested *MAT***a** cells do not use the left arm donor as much, the break still seems to be repaired with the same efficiency, albeit, using the right arm donor instead of the left arm donor (Coïc et al. 2006). This implies that it is the left arm usage, through the activation of the RE, that is cell cycle dependent rather than other aspects of mating type switching such as DNA repair.

The DNA helicase, Chl1 is also important for RE activity but doesn't bind the RE directly (Weiler et al. 1995; Coïc et al. 2006). Chl1 also plays a role in sister chromatid cohesion during S-phase. This implies that sister chromatid cohesion could be an important part of the mechanism behind switching. It is currently unknown how having two chromosomes attached together after replication will affect the switching process.

DNA repair of the broken *MAT* locus

Study of the mating type switching system has been instrumental in understanding the molecular mechanisms behind DNA break repair. After HO cleaves the *MAT* locus, the ends of the break are made single stranded via exonuclease digestion of the 5' ends by MRX (Mre11, RAD 50, Xrs2) (White and Haber 1990; Ivanov et al. 1994). This resection can extend for several kilobases. The single stranded DNA is bound by RPA1 and RAD51 which facilitates homology search and strand invasion at the Z region of the donor locus (White and Haber 1990; Umezu et al. 1998; Miyazaki et al. 2004; Sugawara and Haber 2006; Hicks et al. 2011; Haber 2012a). The highly positioned nucleosomes at the donor locus are transiently remodeled by RAD54 to allow for strand invasion (Hicks et al. 2011). The 3' end of the cut site is used to prime the extension of nascent DNA using the donor as the template. This extension of the DNA strand incorporates the W and X regions from the donor which are homologous to the

upstream W and X regions of the *MAT* locus. Mismatches in the duplex are corrected at this stage through a PMS1-dependent mechanism (Ray et al. 1991; Haber et al. 1993). The nascent strand then dissociates and the W and Z region and anneals to the complementary sequence in the *MAT* locus. The 3' end of the original *MAT* sequence is clipped off by a process that requires RAD51, removing the original Y region from the *MAT* locus (Fishmanlobell and Haber 1992). This 3' end is then used to prime DNA extension using the nascent strand as a template. Finally, the gene conversion is complete when nicks in both strands are ligated together. This mechanism only changes the sequences at the *MAT* locus and doesn't affect the donor sequences. RAD15, RAD52, RAD51, RAD54, RAD55, and RAD57 are also important for repairing the break at *MAT* (Sugawara et al. 1995; Ivanov et al. 1996).

Donor preference

One of the earliest rules that were discovered is that *MAT***a** cells preferentially chose the left donor whereas *MAT*α cells preferentially chose the right donor during switching (Weiler and Broach 1992a). The *HM* loci themselves can be switched in their chromosome positions but donor preference remains the same (Weiler and Broach 1992a). This means that the location of the donor is important for its cell type specific usage, whereas the sequence of the donor is not important. The nucleosomal architecture is very similar at *HML* and *HMR*

and thus is likely not the discriminating factor between them. This lead us and others to hypothesize that chromosome conformation plays a role in donor preference.

The leftmost Mcm1p/*MAT*α2p (DPS1, Donor Preference Site 1) operator in the RE plays a role in donor preference in both *MAT***a** cells and *MAT*α cells. When the *MAT*α2p binding site is mutated, it has no effect on *HML* usage in *MAT***a** cells (Wu et al. 1998a). However, *HML* usage significantly increases in *MAT*α cells, suggesting that the role of the *MATα2p* binding site is to shut down the RE in *MAT*α cells (Wu et al. 1998a). Mutation of the MCM1p sites reduces the *HML* usage to background levels in *MAT***a** cells but doesn't affect *HML* usage in *MAT*α cells (Wu et al. 1998a). The binding of *MAT*α2p to the RE in *MAT*α cells phases the nucleosomes which occlude the Fkh1 binding sites in other parts of the RE (Wu et al. 1998a). Mutation of the *MAT*α2 binding site causes a de-phasing of the nucleosomes in alpha cells. This is presumably due to the binding of MCM1p which is not affected by the mutation. Interestingly, mutation of the MCM1 binding site causes the nucleosomes to become phased in *MAT***a** cells (Wu et al. 1998a) but doesn't affect the phasing of nucleosomes in *MAT*α cells, presumably because *MAT*α2 can still bind to the operator (Szeto et al. 1997a). These data suggest that Mcm1p and Mcm1p/*MAT*α2p have antagonistic roles in establishing the chromatin structure at the RE, and that the default chromatin state of the RE is a closed or phased conformation of the nucleosomes.

The right most part of the RE also plays a role in donor preference but to a lesser extent. This part of the RE which is ~1,800 bp contains the second MCM1p-*MATα2p* operator (DPS2, Donor Preference Site 2) and ARS304. Deleting the 31bp DPS2 site did not change the amount of left arm donor usage in either mating type (Szeto et al. 1997a). However, deleting large regions (~300 – 1,800 bp) around DPS2 reduces left arm usage by 20-50% in *MAT***a** cells (Wu and Haber 1996b; Szeto et al. 1997a). Like DPS1 in the left part of the RE, DPS2 seems to act as a switch that activates the right part of the RE and the sequences around DPS2 seem to be actively involved in donor preference. This region has been less extensively studied than the left part of the RE and an in depth analysis of the factors that bind to this region is necessary to fully understand its role.

Interestingly, the amount of homology in the Z region of the donor can change how competitive one donor is with the other donor (Coic et al. 2011). For instance, in an RE deletion strain, the usage of *HML* can be increased by increasing the amount of homology at the Z region in *HML* (Coic et al. 2011). This phenomenon is employed naturally in the switching system, because *HML* has slightly more homology to the *MAT* locus in the Z region than does *HMR* (Haber 2012a). Increased homology of regions upstream of the donor may contribute to the efficiency of homology searching when the nascent DNA copied from *HML* tries to anneal back to the *MAT* locus. Thus, the additional homology in the W region that is specific to *HML* could further increase left arm preference.

Donor preference depends on chromosomal position

Several sets of experiments illustrate that donor preference is strongly related to the position of each component of the mating type switching cassette (Wu and Haber 1995c; Wu et al. 1996). Interestingly, the RE's influence over donor preference is less efficient when the RE is moved further away from the donor it should be "enhancing" (Szeto et al. 1997a; Coic et al. 2006). The RE effect is also strongest when it is on the same chromosome arm as the donor (Coic et al. 2006). Experiments in which the left donor was moved closer towards the centromere revealed decreased *HML* usage (Wu and Haber 1995a). It is hard to distinguish whether these effects are due to confinement of these loci to a smaller volume around the centromeric region and thus the loci cannot sample enough space to drive efficient repair of *MAT*, or whether the effect of the RE deceases with the distance from the donor (Duan et al. 2010a; Tjong et al. 2012). This alone may decrease the probability of the donor finding *MAT* and therefore decrease repair of the break using that donor. It has been suggested that the RE acts in a catching mechanism through the action of Fkh1p binding to the RE, and the FHA domain of Fkh1p binding to phospho-threonine residues at the break at *MAT* (Li et al. 2012). This mechanism may make the RE "stick" to sites of DNA repair. This mechanism would increase the probability that the donor locus would interact with the broken *MAT* locus. Therefore, if the donor is

too far away from the RE then it will not be able to capitalize on this gain in contact probability.

When *MAT* is moved to the left arm, the RE is not needed for *HML* to be used as a donor in **a** cells (Coic, Richard et al. 2006). This is because *HML* and *MAT* now have a much higher contact probability than they had when *MAT* was on the right arm, but when the RE is moved to the right arm in these strains, there is an increase in right arm usage (Coic, Richard et al. 2006). This indicates that in *MAT***a** cells, the donor that is on the same arm as the RE is chosen as long as it is close to the RE. The RE can also enhance the usage of the donor near it when *MAT* is on a different chromosome or when the RE and the donor are both located on a different chromosome from *MAT* (Houston et al. 2004)(Coic, Richard et al. 2006).

Transcription alters chromatin architecture at the re

DPS1 and DPS2 sites have been shown in reporter assays to be modest activators due to the action of Mcm1 alone (Szeto, Fafalios et al. 1997). However these sites are very strong repressors when the Mcm1p/*MATα2p* complex binds (Szeto, Fafalios et al. 1997)(Zhong et al. 1999).

In *MAT***a**, cells there are several non-protein coding transcripts produced downstream of DPS1 (**Fig. 1.2**), and a small non-coding transcript is produced downstream of DPS2 (Szeto, Fafalios et al. 1997) (Ercan, Reese et al. 2005).

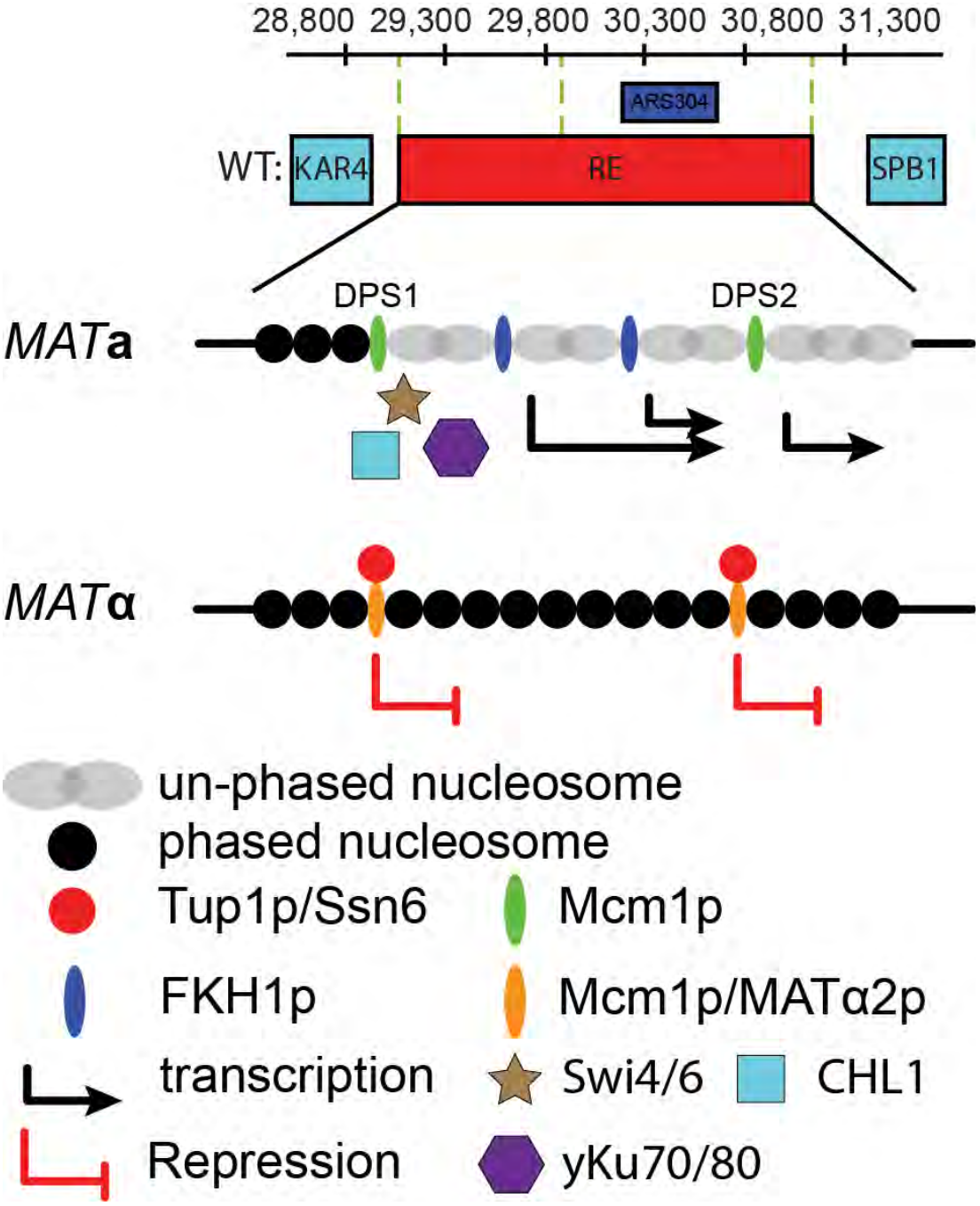


Figure 1.2. Schematic of the Recombination Enhancer. The RE is located on the left arm of Chr. III between KAR4 and SPB1. Action of Mcm1p and Swi4/6 disrupts the phasing of nucleosomes in *MAT***a** cells. Transcripts emanate from the RE in *MAT***a** cells. yKu70/80, Swi4/6, and CHL1 play a role in RE function in the G1 stage of the cell cycle whereas Fkh1 binds to several locations of the RE in G2. In *MAT*α cells, Mcm1p/*MATα2p* bind to the RE and repress transcription.

This transcription is necessary for left arm usage in *MAT***a** cells (Ercan, Reese et al. 2005). The Mcm1p binding site is not essential for transcription at the RE nor left arm usage when another transcription activator is inserted in the RE (Ercan, Reese et al. 2005). When both the DPS1 and DPS2 operators are deleted, there is transcription in the RE in both *MAT***a** and *MAT***α** cells although the transcripts are longer and only produced downstream of where the DPS1 site would be (Szeto, Fafalios et al. 1997). It is possible that this is read-through transcription from the KAR4 locus upstream of the RE but this has not been shown. Transcription occurs in *MAT***α** cells in this instance because the binding sites for the *MATα2p* repressor are no longer available.

Why is transcription important for RE function? One might think that since Fkh1p, which binds to several places in RE-left, is a transcription factor, then its binding initiates transcription. However, the opposite is true; Fkh1p binding is dependent on active transcription (Ercan et al. 2005). This transcription results in unphased nucleosomes in *MAT***a** cells (Weiss and Simpson 1997; Ercan and Simpson 2004; Ercan et al. 2005). These data suggest that transcription might open the chromatin at the RE to allow for binding of other factors. In *MAT***α** cells, the RE is repressed by the action of Mcm1p/*MATα2p*. This results in highly positioned nucleosomes that cover up Fkh1 binding sites, preventing Fkh1 binding (Weiss and Simpson 1997; Ercan et al. 2005).

These data show that transcription seems to act as the key that activates the RE in *MAT***a** for several reasons. First, no role has yet been described for the

transcripts at the RE. Second, transcription in general, and not a specific transcript from the RE is important for RE function (Ercan et al. 2005). Finally, the lack of transcription results in phased nucleosomes that makes the chromatin of the RE refractory to DNA binding proteins (Ercan et al. 2005).

RE function is not specific to switching and may involve a general mechanism

One would presume that the RE, being part of the mating type switching system, would only be functional when the *HML* is needed to repair a break at *MAT*. In fact, the RE enhances recombination in general along the same arm that the RE is on in *MAT***a** cells. Studies measuring the spontaneous recombination between Leu2 alleles showed higher numbers of recombinates in *MAT***a** cells with a functional RE (Wu and Haber 1995a; Sun et al. 2002b; Houston et al. 2004). The rules that govern this process seem to be the same as those that govern the RE's ability to enhance recombination between the left arm donor and *MAT*. This lack of specificity indicates that the RE might work through a general mechanism that is not directly linked to mating type switching. One such general mechanism that has been shown to play a role in DNA break repair is nuclear organization (Zhang et al. 2012a). DNA breaks have been shown to be preferentially repaired with DNA that is in the same nuclear compartment or "neighborhood", even though these sections of DNA could be far away from each

other in the linear genome (Zhang et al. 2012a). Since nuclear topology has been shown to be an integrated part of DNA break repair in other systems, I hypothesize that it could play an important role in mating type switching as well.

The recombination enhancer requires yKu70/80

yKu70 and yKu80 play a role in several biological processes in budding yeast including telomere stability, telomere positioning, and DNA repair. Binding of yKu80 to the left portion of the RE requires an intact MCM1p binding site at DPS1 (Ruan et al. 2005). However, this may be due to MCM1p's role in opening the chromatin of the RE and not due to a direct contact with MCM1p. Deletion of yKu80 in *MAT***a** cells alters the chromatin structure in the RE to the *MAT*α, "closed" conformation (Ruan et al. 2005). This results in a decrease of left arm usage during switching (Ruan et al. 2005). yKu70 is also important for left arm usage. Yku70 binds to the RE before break induction and then spreads down the left arm toward *HML*. By one hour after break induction yKu70 is exclusively located at *HML* but not at *HMR*. Deletion of yKu70 also makes *HML* adopt a more peripheral localization in the nucleus (Bystricky et al. 2009a; Miele et al. 2009).

Another group observed that yKu70 and yKu80 bind to *HML* and *HMR* in both *MAT***a** and *MAT*α cells, but they bind more to *HML* in *MAT***a** cells and more to *HMR* in *MAT*α cells. yKu70 and yKu80 are also involved in the silencing of the

HM loci (Patterson and Fox 2008). Binding or yKu70 to *HML* is dependent on the RE and SIR4 (Bystricky et al. 2009a). Possibly, yKu70 acts as a messenger between the RE and *HML* during switching, and after break induction yKu70 travels from the RE to *HML* and assist in releasing *HML* from the nuclear periphery.

The RE modulates nuclear organization of switching components

Experiments tracking the mobility and nuclear positioning of the mating type switching loci have revealed that *HML* is less constrained in **a** cells than in *MAT***α** cells (Bressan et al. 2004b). However, *HML* has a slight enrichment for localization at the periphery in both mating types (Bystricky et al. 2009a). When the RE is deleted in *MAT***a** cells, *HML* adopts a more peripheral localization (Bystricky et al. 2009a). This corresponds with *HML* becoming more constrained in its movement (Bressan et al. 2004b). There seems to be little or no difference between the mobility of *HMR* in *MAT***a** and *MAT***α** cells, but deleting the RE only slightly reduces the mobility of *HMR* in *MAT***a** cells. This is likely an indirect effect due to the fact that *HML* and *HMR* physically interact with one another through their silencers (Miele et al. 2009). Thus, a change in the left arm's mobility in *MAT***a** cells will also cause a change in the mobility of *HMR*. These data suggest that the left arm of Chr. III may become tethered to the nuclear periphery in the absence of the RE. Tethering of the ends of Chr. III is mediated

by the telomeres (Bystricky et al. 2009a). This is consistent with the proposed function of yKu70/yKu80 as a messenger from the RE to the end of the left arm, given its role in localizing telomeres to the periphery. The HO induced double stranded break at the *MAT* locus during switching also becomes confined to the nuclear periphery when a donor is absent (Oza et al. 2009).

Yeast nuclear organization

Interestingly, budding yeast chromosomes, which are 100's of kb – 1's Mb long, have a very different chromatin topology than mammalian chromosomes. Studies employing chromosome conformation capture techniques, live cell imaging, and polymer simulations have revealed global and some local features of the nuclear organization in budding yeast. For instance, very strong interactions between pairs of centromeres have been observed (Berger et al. 2008; Duan et al. 2010a). This is due to a clustering of centromeres at one pole of the nucleus adjacent to the Spindle Pole Body (SPB), which was initially observed by microscopy (Jin et al. 1998a). Yeast nuclei display prominent centromere clustering that is reduced in non-dividing cells and in meiotic prophase (Jin et al. 1998a; Jin et al. 2000a). Centromere clustering is also a major determinant of yeast interphase nuclear organization (Jin et al. 2000a). Strong interactions between pairs of telomeres located on separate chromosomes have also been observed (Berger et al. 2008; Duan et al. 2010a).

This phenomenon is consistent with observations that telomeres are tethered to the nuclear periphery (Gotta et al. 1996). The clustering of telomeres coincides with a co-localization with Rap1, Sir3, and Sir4 proteins in wild-type Saccharomyces cerevisiae (Gotta et al. 1996; Trelles-Sticken et al. 2000). Meiotic telomere protein Ndj1p is required for meiosis-specific telomere distribution, bouquet formation and efficient homologue pairing (Trelles-Sticken et al. 2000). Limiting the mobility of the telomeres to the 2D periphery may drive their co-localization by restricting their possible locations.

The structure of Chr. XII is very interesting due to the rDNA array located on the right arm of Chr. XII. Only two copies of the rDNA locus are included in the reference genome assembly. In reality there are ~150 copies of the rDNA per cell (Kobayashi et al. 1998). This massive array of rDNA repeats forms a crescent shaped nucleolus that is located at the pole opposite the SPB (Berger et al. 2008). Therefore, unique constraints are placed on the two portions of the right arm of Chr. XII. For the portion of the right arm that is upstream of the rDNA array, it is both tethered to the SPB and the nucleolus, making it the only portion of a chromosome that is tethered to both poles of the nucleus. The portion of the right arm that is downstream of the rDNA array is the only chromosome section that is anchored at the opposite pole from the SPB, and thus it projects into the nucleus from the opposite pole than all of the other chromosome arms (Duan et al. 2010a).

It has been observed that there is a difference in the behavior of short chromosome arms compared to long chromosome arms (Tjong et al. 2012). Short arms seem to occupy a volume very close to the SPB, whereas long arms extend out more. Polymer simulations have suggested that this is an effect of volume exclusion near the SPB (Tjong et al. 2012).

Some aspects of yeast nuclear organization are not constitutive, but rather are acquired depending on the specific conditions of the cell. It has been observed that SAGA-dependent genes are confined to the nuclear periphery when activated (Cabal et al. 2006).

Nuclear organization in higher eukaryotes

Studies of chromatin conformation, using 3C-based techniques, have revealed many aspects of chromatin structure. It is clear that chromatin is not randomly dispersed in the eukaryotic nucleus. In higher eukaryotes (e.g. mouse, human, and flies) the genome is organized hierarchically. The highest level of organization is chromosome territories that are on the order of 10's - 100's of Mb in size. Chromosome territories are formed because the chromosomes only rarely intermingle with other chromosomes. The next level of hierarchy, chromosome compartments, are formed by the association of large (1's -10's Mb) regions of chromatin that have similar properties (active and inactive) and are spatially separated from each other in the genome (Lieberman-Aiden et al.

2009a; Zhang et al. 2012a). Compartments are further subdivided into Topologically Associating Domains (TADs). TADs are regions of 100's kb – 1'sMb where chromatin preferentially interacts with itself but not with neighboring regions (Dixon et al. 2012a; Nora et al. 2012). These hierarchical levels of chromatin structure play a role in many biological functions such as gene expression (Bau et al. 2011; Sanyal et al. 2012; Jin et al. 2013; Symmons et al. 2014) and DNA repair (Zhang et al. 2012a). Defects in this organization have also been implicated in human pathology, e.g. in the premature aging disease Hutchinson-Gilford Progeria Syndrome (McCord et al. 2013).

The properties of chromatin organization at the sub-TAD length scale in mammals, but higher order than nucleosomal DNA (10's bp – 1's Kb), are less well characterized. It is costly to perform comprehensive chromatin interaction studies at the resolution of single restriction fragments in organisms with large genomes because the number of possible pairwise chromatin interactions scales with the square of the number of restriction fragments in a genome. Organisms with small genomes present a unique opportunity to study the conformation of chromatin at high resolution in a cost-effective manner.

Chromosome conformation in other organisms

Chromatin conformation capture studies in other organisms are yielding interesting insights in to the biology of the genomes of non-mammalian

organisms. For instance, experiments of the biology of *Caulobacter crescentus* swarmer cell chromosomes revealed that the chromosome is not merely a circle but in fact is an ellipse with a helical twist (Umbarger et al. 2011). The termini of this chromosome corresponds to genomic loci which are involved in chromosome segregation and replication. High resolution analysis of the chromosome in this organism revealed that distinct Chromatin Interaction Domains (CIDs) are present (Le et al. 2013). These domains are analogous to TADs which show an enrichment of interactions between loci within the CID but a depletion of interaction with loci in other neighboring CIDs. These CIDs are a byproduct of transcription elongation and are due to the formation of plectonemes formed from supercoiling of the DNA as transcription bubbles move through these domains. These experiments also revealed that SMC (Structural Maintenance of Chromosomes) proteins, a protein family found in many organisms, play a role in maintaining the co-linearity of the chromosome arms which result in the ellipse structure of the chromosome. This shows that chromosome conformation is such a basic phenomenon that even the simplest organisms utilize it as part of their biology.

Chromosome conformation capture has also been employed to analyze the plant species *Arabidopsis thaliana* (Grob et al. 2013; Feng et al. 2014; Grob et al. 2014). The gross organization of the chromosomes, in this species, indicates that pericentromeric regions interact with one another very strongly but are depleted for interactions with the arms of the chromosome. Additionally,

telomeres interact with one another very strongly. These features are reminiscent of the Rabl configuration of chromosomes found in both budding yeast and fission yeast (Duan et al. 2010a; Mizuguchi et al. 2014). However, there are some distinct differences in the organization of *Arabidopsis* chromosomes. First the pericentromeric regions and other heterochromatic loci are more condensed than other portions of the genome. Second, there is an enrichment of interaction between heterochromatic loci both on the same chromosome and on different chromosomes at Interactive Heterochromatic Islands (IHIs). Regions with active histone marks also interact with one another. This sort of partitioning of active and inactive regions of the genome is reminiscent of chromatin compartments observed in human and mouse (Lieberman-Aiden et al. 2009a; Zhang et al. 2012a; McCord et al. 2013), albeit on a smaller length scale. Although there are many structural properties in common between *Arabidopsis* and other organisms, there doesn't seem to be a consecutive array of self-interacting domains similar to Topologically Associating Domains (TADs).

In Drosophila, there also seems to be an enrichment for centromere-centromere interactions on the autosomes. Interactions between some telomeres is also evident. The Drosophila genome is also partitioned into active and inactive regions similar to compartments, except the inactive regions are repressed by Polycomb group (PcG) proteins (Sexton et al. 2012).

In Drosophila, Arabidopsis, and S. pombe, long-range looping interactions between specific loci are observed. These interactions are distinct from the interaction patterns that are observed for compartment structures (Sexton et al. 2012; Grob et al. 2014; Mizuguchi et al. 2014). These interactions are probably not present at all times in every cell in the population since their signal intensity is not very high compared to neighboring interactions which interact very frequently with one another. This suggests that there are either temporal dynamics of these interactions or heterogeneity in the cell population.

The role of chromatin conformation in gene regulation

Chromatin loops have been observed for several decades using electron microscopy (Griffith et al. 1986), light microscopy (Williamson et al. 2012), RNA TRAP (Carter et al. 2002), and Chromosome conformation capture techniques (Tolhuis et al. 2002; Sanyal et al. 2012). One of the classic and most studied examples of looping between promoters and regulatory elements is the looping between the Locus Control Region (LCR) and the beta-globin genes in both mice and humans (Carter et al. 2002; Tolhuis et al. 2002; Palstra et al. 2003). This looping has been observed to be cell type dependent, in that the loops are made in cells expressing the beta-globin genes and are not present in cells that do not express these genes (Tolhuis et al. 2002; Palstra et al. 2003).

Chromatin looping interactions have also been observed in the T_H2 cytokine locus (Spilianakis and Flavell 2004), The mouse alpha-globin locus (Vernimmen et al. 2007; Bau et al. 2011), and the CFTR locus (Gheldof et al. 2010). Large scale analysis has found that there are many chromatin loops between regulatory elements and that these loops are largely cell type specific (Sanyal et al. 2012). Chromatin loops have also been shown to be important in the developmental regulation of genes. Chromatin interactions between regulator elements and promoters have been observed to change during differentiation (Vernimmen et al. 2007; Andrey et al. 2013).

Anchoring of the genome to the nuclear periphery is also important for the regulation of gene expression. There are certain regions of chromosomes in human that interact with nuclear matrix proteins such as Lamin B1, Lamin A/C, and Emerin (Guelen et al. 2008; Zullo et al. 2012; Kind et al. 2013; McCord et al. 2013). These Lamin Associated Domains (LADs) are depleted for genes and enriched for the repressive mark, H3K27me3. LADs correlate very well with the inactive spatial compartment of the genome (McCord et al. 2013). This leads to the model that LADs are inactive heterochromatin that is sequestered to the nuclear periphery. This could potentially be the mechanism the partitions the genome spatially and functionally into the two compartments. Upon differentiation, some loci are released from the nuclear periphery and become activated (Zullo et al. 2012).

Chromosome conformation plays a role in partitioning co-regulated genes. Genes in the HOXD cluster, in mouse forelimbs, have been shown to reside in different Topologically Associating Domains (TAD) based on the stage of differentiation and the resulting cell type (Andrey et al. 2013). Chromatin topology also plays a role in the organization of hormone-induced genes. Genes induced or repressed by Progestin tend to cluster in TADs and genes that are induced or repressed by Estradiol also tend to cluster in to TADs (Le Dily et al. 2014). This clustering extends further, in that the Progestin and Estradiol responsive genes share similar TADs. This suggests that not only are genes responsive to a single hormone organized together but, perhaps, hormone responsive genes in general are organized together in an additional level of hierarchy (Le Dily et al. 2014). In other words genes may be organized into structural hierarchies based on their biological function.

Chromosome conformation and human disease

The evidence presented above illustrates that nuclear organization is involved in many biological processes. It stands to reason then, that defects in chromosome conformation will disrupt biological processes and potentially result in diseases. This is indeed what is observed. As mentioned above, repressive chromatin associates with the nuclear matrix in what are called LADs. Patients with Hutchinson-Gilford progeria syndrome have a mutation in Lamin A which

disrupts the association of the chromatin to the nuclear matrix (McCord et al. 2013). The overall nuclear morphology becomes less spherical and more irregularly shaped. This alone could disrupt chromatin interactions. These patients also experience a loss of the repressive mark H3k27me3 in the LADs which alters expression profiles in these patients. Genome structure may also influence the translocations that lead to cancer.

Translocations, e.g. in cancer cells, have been observed to happen most frequently between chromatin that is on the same chromosome and even more so between segments of chromatin that are in the same spatial compartment and on the same chromosome (Zhang et al. 2012a). Thus, the spatial organization of chromosomes can guide the occurrence of chromosomal alterations in cancer.

Methods for studying nuclear organization and chromatin structure

3C based techniques

The Chromosome Conformation Capture (3C) technique and its derivatives measure the population-averaged frequency at which two DNA fragments physically associate in three-dimensional (3D) space, based on the propensity for those two locations to become formaldehyde-crosslinked together (**Fig. 1.3**). Once interacting loci are crosslinked, chromatin is solubilized and fragmented, usually using a restriction enzyme (**Fig. 1.3**). Interacting fragments are then ligated together and purified, creating a genomic library of chimeric DNA

molecules (**Fig. 1.3**). The relative abundance of specific chimeras, or ligation products, is related to the probability that those fragments interact in 3D space across the cell population. A 3C library includes a massive variety of ligation products (up to 10^{12} unique pair-wise interactions between 4 Kb fragments in the human genome) and can be analyzed in various ways depending on the goals of the study (**Fig. 1.3**). To evaluate specific interactions, 3C ligation products can be assayed individually using PCR (**Fig. 1.3**). In special variants of the 3C technology (ChIP-loop and ChIA-PET), immunoprecipitation is used to associate the interactions with a particular protein of interest (Horike et al. 2005; Fullwood et al. 2009a). ChIA-PET is used to investigate the role of the protein factor in facilitating genomic contacts. In recent years, new approaches such as 4C, 5C, and Hi-C have been developed to utilize Next Generation Sequencing (NGS) technologies in order to interrogate the 3C ligation product library more comprehensively (**Fig. 1.3**) (Dostie et al. 2006a; Simonis et al. 2006a; Zhao et al. 2006a; Lieberman-Aiden et al. 2009a). Most 3C-based techniques focus on analysis of a set of predetermined loci enabling “one-versus-some” (basic 3C and ChIP-loop), “one-versus-all” (4C), or “many-versus-many” (5C) explorations of the conformation of chromosomal regions of interest (**Fig. 1.3**). On the other hand, Hi-C enables an “all-versus-all” interaction profiling (**Fig. 1.3**).

The type of 3C-based technique that should be employed for a given study depends on the question and scope of the experiment. All of these techniques rely on formaldehyde crosslinking of chromatin and subsequent

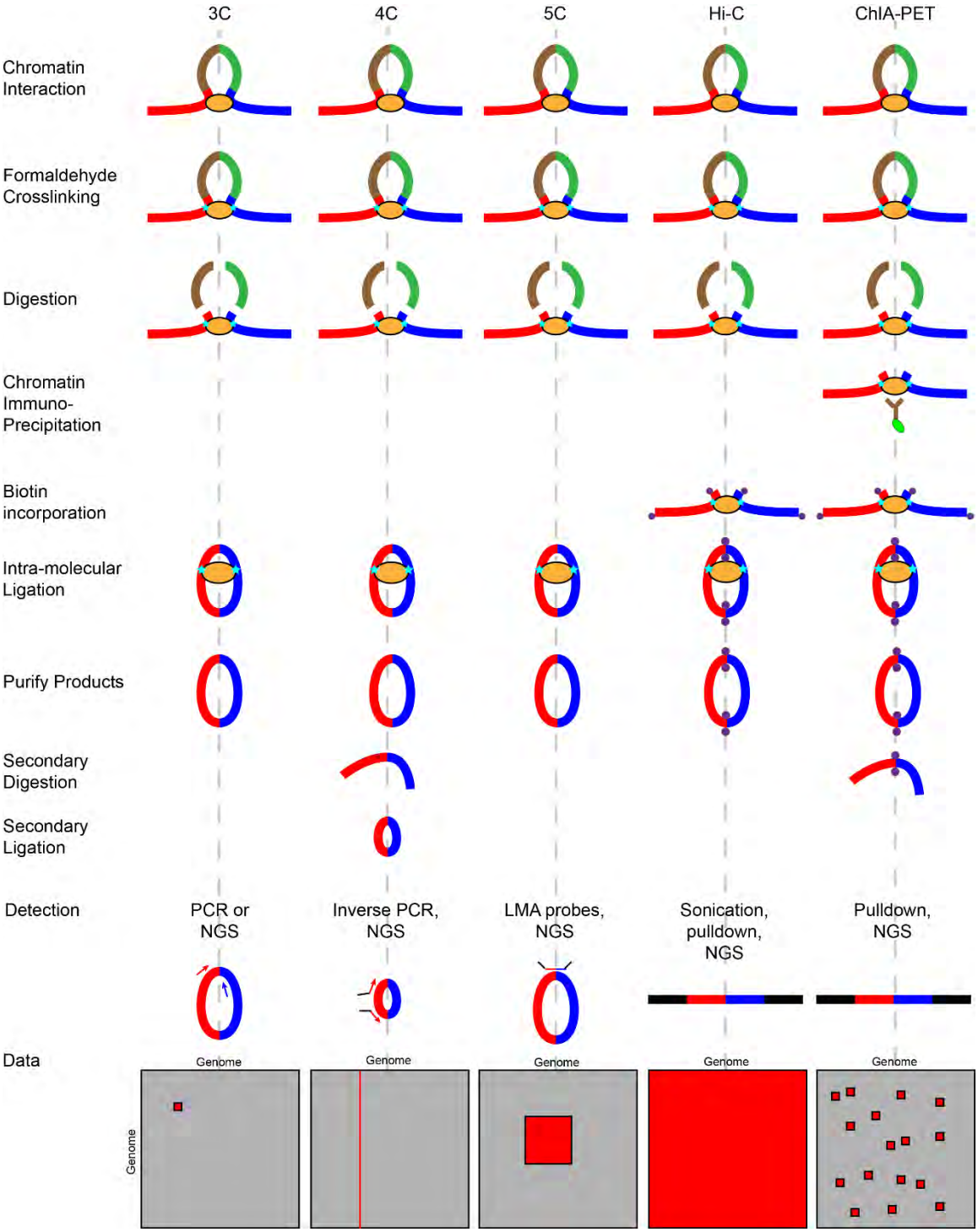


Figure 1.3. Overview of some chromosome conformation techniques. This figure shows the general protocol for several of the chromosome conformation techniques. Red, green, blue, and brown bars represent chromatin restriction

fragments. Yellow oval represents protein complexes which mediate genomic interactions. Cyan stars represent formaldehyde crosslinks. Purple circles represent biotinylated nucleotides. The large boxes at the bottom represent all of the interaction space possible genome-wide and the red boxes represent the type of data that each particular assay yields.

proximity ligation of crosslinked restriction fragments in order to detect chromatin interactions. The only difference is the manner in which the products are detected.

3C-PCR is applicable to studies focusing on only a few (key) regions/loci of interest. Typically, experiments with less than 50 restriction fragments to be analyzed can be readily carried out using classical 3C followed by PCR. 4C is suitable for studies interrogating genome wide interactions of a single locus of interest, e.g. a centromere. For more comprehensive studies of target regions, e.g. whole chromosomes, that are not genome-wide, 5C is more applicable. 5C allows for the high-throughput detection of chromatin interactions in a single reaction (Dostie et al. 2006a). Short probes are used to hybridize to ligation junctions of interest in a 3C library. When two chromatin fragments of interest are ligated to each other in a 3C library, the probes anneal adjacent to one another and are then ligated together with Taq Ligase. This, in effect, makes a "carbon-copy" of the ligation product. These products are then detected with high throughput sequencing of the short ligated probes. The types of questions that can be addressed with 5C depend on how the probes are designed. The probes can be arranged for all restriction fragments along a chromosome which would yield comprehensive information of the spatial distribution of the entire chromosome. This information could be used to generate three-dimensional (3D) models of the chromosome (Bau and Marti-Renom 2011; Bau et al. 2011). The probes could also be designed so that one set hybridizes to restriction

fragments containing regulatory elements and another set hybridizes to fragments overlapping genes. This design would allow for a comprehensive analysis of the spatial interactions between regulatory elements and their genes (Sanyal et al. 2012). Many other variations of these two generic probe designs are possible. 5C experiments typically use 50-5000 probes to detect hundreds to millions of chromatin interactions in parallel.

Recently, ChIA-PET has merged 3C with chromatin immunoprecipitation assays (Fullwood et al. 2009b). In this method, proteins of interest are immunoprecipitated from the crosslinked and digested chromatin. This adds an additional level of complexity in that protein-DNA interactions can be assayed at the same time as chromatin interactions.

When the goal of the experiment is to obtain information regarding the spatial organization of a complete genome, then Hi-C is the technique of choice (Lieberman-Aiden et al. 2009a; Duan et al. 2010a; Umbarger et al. 2011; Dixon et al. 2012a; Sexton et al. 2012; Zhang et al. 2012a; Grob et al. 2013; Le et al. 2013; McCord et al. 2013; Naumova et al. 2013; Feng et al. 2014; Grob et al. 2014; Mizuguchi et al. 2014). In Hi-C, the ends left after restriction digestion are filled in with biotinylated nucleotides prior to ligation. As a result, ligation junctions are labeled with biotin and can be purified in an unbiased manner using streptavidin-coated agarose beads. Purified ligation junctions are then analyzed and quantified using high throughput sequencing to obtain an unbiased genome-wide interaction map.

Microscopy techniques

Microscopy based techniques are also used to study nuclear organization and gross chromosome conformation. All of these techniques provide unique perspectives of nuclear organization. These techniques include electron microscopy, which is used to analyze the architecture of the nucleus in nm scale detail, and fluorescent (light) microscopy, which provides information on the shape and distribution of specific chromosomes and chromosomal loci as well as the co-association of specific loci with sub-nuclear compartments with a resolution of 50-100 nm (Dehghani et al. 2005). The major disadvantage of electron microscopy is its lack of connectivity to the genomic sequence, in that the incredible fine resolution architecture cannot be assigned to a specific location in the genome. Using sequence-specific probes in Fluorescent In Situ Hybridization (FISH) assays, connects nuclear architecture and DNA sequence, but these methods are currently still limited in throughput, allowing analysis of only a few loci simultaneously (Solovei et al. 2002).

Understanding the nuclear context is important to fully understand chromosome structure because the nuclear environment (nuclear matrix, nucleoli, etc.) places constraints on the chromatin. Live and fixed cell microscopy can be used to measure the distribution of loci in the nuclear volume (Cabal et al. 2006; Berger et al. 2008; Bystricky et al. 2009a; Miele et al. 2009).

Knowing where a loci is in the nucleus only provides a static perspective of the biology of that locus. Live cell microscopy can also be used to measure the diffusion coefficient of a locus as well as the amount of space a locus samples or its confinement (Berg 1983; Bressan et al. 2004b; Bystricky et al. 2009a; Miele et al. 2009; Meister et al. 2010). The diffusion coefficient can be thought of as the velocity of the locus being studied and the radius of confinement is how much volume the locus is free to move in. Both of these pieces of information add to the understanding chromatin interactions in ways that cannot be assayed by static techniques.

In this thesis, I have employed the 5C and Hi-C techniques to comprehensively study the conformation of chromosomes in budding yeast. I have found that Chr. III is the only chromosome which has a difference in conformation between the mating types. This chromosome can be thought of as the sex chromosome in yeast because it contains the mating type switching cassette. Deletion of the RE, a cis regulator element involved in mating type switching, ablates these differences in conformation between the mating types. Both *MAT***a** and *MAT*α conformations are effected when the RE is deleted. Further dissection of the RE revealed that the left portion, that is responsible for donor preference in *MAT***a** cells does not affect the structure in these cells. However, this portion does have a minor structural role in *MAT*α cells. The right portion of the RE is responsible for the mating type specific conformations and affects the structure in both mating types. This analysis has revealed that the RE

is a composite element with the left portion playing a catalytic role in *MAT***a** cells during switching and the right portion playing a structural role in both mating types. This work has broadened our understanding of the mechanisms that establish chromosome structure.

Chapter II

Hi-C: A Comprehensive Technique to Capture the Conformation of Genomes

Contributors

The text and figures from this chapter are from: Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. 2012. Hi–C: A comprehensive technique to capture the conformation of genomes. *Methods* **58**: 268-276. The authors that contributed to this work in the order of their contribution are: Jon-Matthew Belton, Rachel Patton McCord, Johan Harmen Gibcus, Natalia Naumova, Ye Zhan, Job Dekker

Abstract

We describe a method, Hi-C, to comprehensively detect chromatin interactions in the mammalian nucleus. This method is based on Chromosome Conformation Capture, in that chromatin is crosslinked with formaldehyde, then digested, and re-ligated in such a way that only DNA fragments that are covalently linked together form ligation products. The ligation products contain the information of not only where they originated from in the genomic sequence but also where they reside, physically, in the 3D organization of the genome. In Hi-C, a biotin-labeled nucleotide is incorporated at the ligation junction, enabling selective purification of chimeric DNA ligation junctions followed by deep sequencing. The compatibility of Hi-C with next generation sequencing platforms makes it possible to detect chromatin interactions on an unprecedented scale. This advance gives Hi-C the power to both explore the biophysical properties of chromatin as well as the implications of chromatin structure for the biological functions of the nucleus. A massively parallel survey of chromatin interaction provides the previously missing dimension of spatial context to other genomic studies. This spatial context will provide a new perspective to studies of chromatin and its role in genome regulation in normal conditions and in disease.

Introduction

The long DNA strands of every cell's genome are packaged into chromatin in a very confined nuclear volume (Cremer and Cremer 2010). The organization of the chromatin in the nucleus is extremely relevant to biological function at the

gene level as well as the global nuclear level. The study of the packaging and organization of chromatin in the nucleus will shed light on the spatial aspects of gene regulation, chromosome morphogenesis, and genome stability and transmission. It will also enhance the understanding of the biophysics of chromatin, and further enable the investigation of pathologies related to genome instability or nuclear morphology.

Several techniques are now available to observe the spatial organization of chromatin. These techniques can be broadly classified into microscopic and molecular assays. Electron microscopy is used to analyze the architecture of the nucleus in nm scale detail, while fluorescent (light) microscopy provides information on the shape and distribution of specific chromosomes and chromosomal loci as well as the co-association of specific loci with sub-nuclear compartments with a resolution of 50-100 nm (Dehghani et al. 2005). The major disadvantage of electron microscopy is its lack of connectivity to the genomic sequence, in that the incredible fine resolution architecture cannot be assigned to a specific location in the genome. Using sequence specific probes in Fluorescent In Situ Hybridization (FISH) assays allows one to connect nuclear architecture and DNA sequence, but these methods are currently still limited in throughput, allowing analysis of only a few loci simultaneously (Solovei et al. 2002). In molecular assays based on Chromosome Conformation Capture (3C) the genomic sequence itself is the output, completely connecting chromosome structure and the genomic sequence (Dekker et al. 2002a).

The 3C technique and its derivatives measure the population-averaged frequency at which two DNA fragments physically associate in three-dimensional (3D) space, based on the propensity for those two locations to become formaldehyde-crosslinked together (**Fig. 2.1a**). Once interacting loci are crosslinked, chromatin is solubilized and fragmented, usually using a restriction enzyme (**Fig. 2.1b**). Interacting fragments are then ligated together and purified, creating a genomic library of chimeric DNA molecules (**Fig. 2.1c**). The relative abundance of specific chimeras, or ligation products, is related to the probability that those fragments interact in 3D space across the cell population. A 3C library includes a massive variety of ligation products (up to 10^{12} unique pair-wide interactions between 4 Kb fragments in the human genome) and can be analyzed in various ways depending on the goals of the study. To evaluate specific interactions, 3C ligation products can be assayed individually using PCR. In special variants of the 3C technology (ChIP-loop and ChIA-PET), immunoprecipitation is used to associate the interactions with a particular protein of interest (Horike et al. 2005; Fullwood et al. 2009a). This is used to investigate the role of the protein factor in facilitating genomic contacts. In recent years, new approaches such as 4C, 5C, and Hi-C have been developed to utilize Next Generation Sequencing (NGS) technologies in order to interrogate the 3C ligation product library more comprehensively (**fig. 2.1D and 2.1E**) (Dostie et al. 2006b; Simonis et al. 2006a; Zhao et al. 2006a; Lieberman-Aiden et al. 2009b). Most 3C-based techniques focus on analysis of a set of predetermined loci

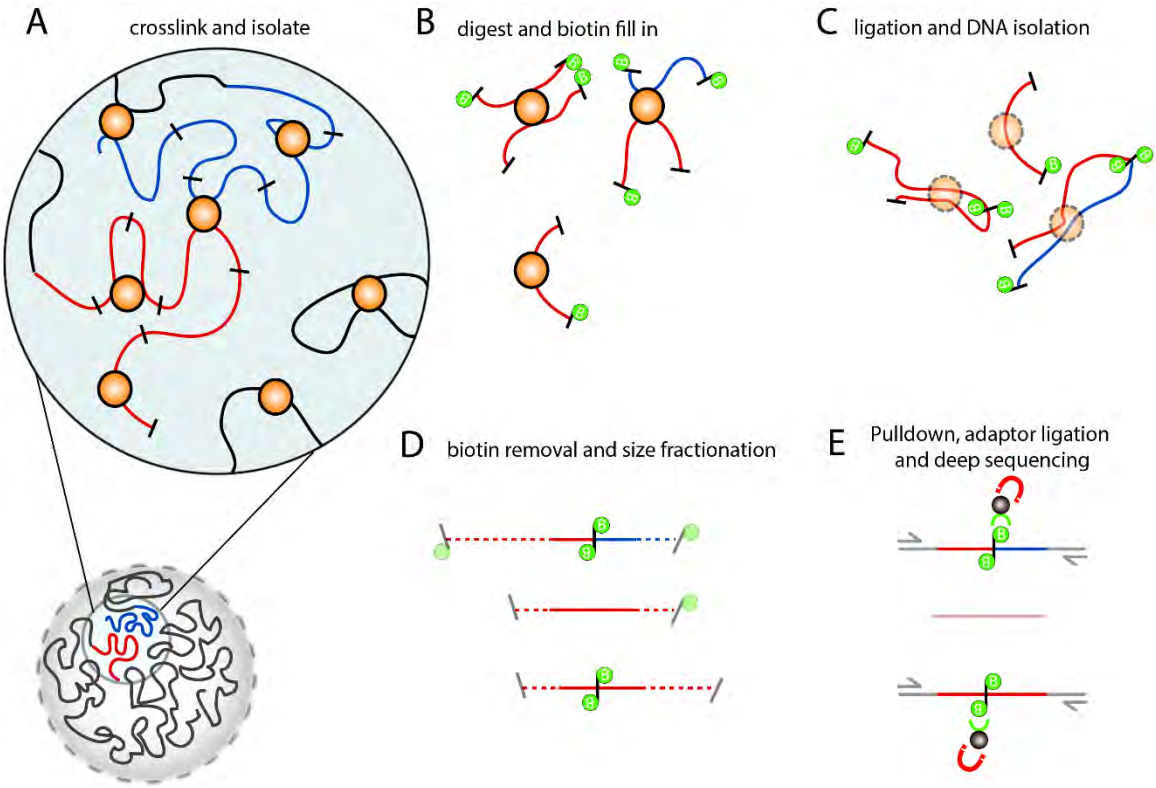


Figure 2.1. Overview of Hi-C technology. A) Hi-C detects chromatin interaction both within and between chromosomes by covalently crosslinking protein/DNA complexes with formaldehyde. **B)** The chromatin is digested with a restriction enzyme and the ends are marked with a biotinylated nucleotide. **C)** The DNA in the crosslinked complexes is ligated to form chimeric DNA molecules. **D)** Biotin is removed from unligated ends of linear fragments and the molecules are fragmented to reduce their overall size. **E)** Molecules with internal biotin incorporation are pulled down with streptavidin coated magnetic beads and modified for deep sequencing. Quantitation of chromatin interactions is achieved through massively parallel deep sequencing.

enabling “one-versus-some” (basic 3C and ChIP-loop), “one-versus-all” (4C), or “many-versus-many” (5C) explorations of the conformation of chromosomal regions of interest. On the other hand, Hi-C enables an “all-versus-all” interaction profiling.

In this paper we describe Hi-C. In Hi-C, all genomic fragments are labeled with a biotinylated nucleotide before ligation, thereby marking ligation junctions. These junctions can then be purified efficiently by streptavidin-coated magnetic beads, enriching the library for ligation products that can be detected by NGS. The comprehensive chromatin interaction data that can be obtained by direct sequencing of a Hi-C library provides immense statistical power for analyses of genome organization at kb resolution. These analyses can reveal overall genome structure and biophysical properties of chromatin as well as more specific long-range contacts between distant genomic elements such as genes and regulatory elements. Further, combining Hi-C data with other datasets, including gene expression profiles and genome-wide maps of chromatin modifications, will allow placing sets of loci in 3D context. This will lead to new insights into the functional roles of chromatin conformation in genome regulation and stability, both in normal cells and in disease states.

1. The Hi-C Method

Cell Culture and Crosslinking of Chromatin.

The combinatorial possibilities of all pair-wise interactions between genomic restriction fragments that can be observed with Hi-C is immense and scales exponentially with genome size. Many of these interaction possibilities may be occurring in the cell population, but, in the experiment, only two interactions for a given fragment can be detected from a single cell. Therefore, it is of critical importance to analyze sufficient numbers of cells to ensure high complexity of the resulting ligation product library. The complexity of this library will ultimately determine the resolution and sensitivity of chromatin interaction datasets. In our hands, a sample of twenty five million cells will typically be enough to produce a sufficiently complex Hi-C library for analysis of global spatial organization in a mammalian genome. For the study of precious samples, such as primary tissue, fewer cells (in the range of 1-5 million) may still generate a successful Hi-C library, but the data from such experiments should be examined for signs of low complexity as discussed in Section 3.2.

Cells are fixed in 1% final concentration of formaldehyde in the relevant culture media. After crosslinking the remaining formaldehyde is sequestered with an excess of glycine. Then the cells are harvested by centrifugation and can be flash frozen and stored at - 80°C for future use.

Standardization of crosslinking conditions is the cornerstone of 3C-based techniques since the functional read-out of this technique is the frequency at

which two genomic restriction fragments are crosslinked to one another. One source of variation in crosslinking is the presence of serum in culture media. Serum, containing a high concentration of protein, can alter the effective concentration of formaldehyde by binding to and sequestering formaldehyde in the culture media. In cases where serum is used for culturing it should be omitted at the time of crosslinking. Special considerations must also be made when making Hi-C libraries from adherent cells. Adherent cells bind to a surface by molecular mechanisms that involve the cytoskeleton, which are known to play a role in the location and shape of the nucleus in the cytoplasm. This linkage between nuclear and cellular morphology suggests that crosslinking of adherent cells while they are still attached to the culturing surface is necessary to preserve the global nuclear organization. Once crosslinked, adherent cells can be scraped from the culture plate using disposable cell scrapers and aliquoted in the same way as suspension cells.

Cell Lysis and Chromatin Digestion.

One pellet of 25 million cells is lysed using a Dounce homogenizer in the presence of cold hypotonic buffer, supplemented with protease inhibitors and a mild non-ionic detergent (IGEPAL CA-630). The presence of a protease inhibitor cocktail during lysis and lysis on ice helps protect the crosslinked chromatin complexes from endogenous proteases. The lysate is washed twice with 1X

restriction enzyme buffer and then resuspended in 1X restriction enzyme buffer. The chromatin is solubilized with dilute SDS and incubation at 65°C for 10 min. This solubilization removes non-crosslinked proteins and opens the chromatin, making it accessible for restriction endonuclease cleavage. Incubation at 65°C can reverse formaldehyde crosslinks, making it important to minimize the incubation time and to place the sample on ice immediately following incubation. Triton X-100, a non-ionic surfactant, is used to quench the SDS, preventing enzyme denaturation in subsequent steps. The now accessible chromatin is then digested with a type II restriction endonuclease, e.g. HindIII, at 37°C overnight while rotating. Any restriction enzyme that produces a 5' overhang could be used to produce a Hi-C library, but the subsequent quality controls should be adjusted accordingly. To avoid star activity (relaxed specificity of cleavage site recognition) of HindIII, the restriction enzyme concentration is minimized in favor of longer incubation. The digestion percentage can be approximated by performing a PCR spanning a specific genomic restriction site both with and without endonuclease treatment. Loss of amplicon signal after endonuclease treatment correlates with digestion efficiency.

Biotin Marking of DNA Ends and Blunt End Ligation.

The HindIII enzyme recognizes the sequence: 5' – AAGCTT – 3' and cleaves the DNA, leaving a 5' overhang of: 5' – AGCT – 3'. This cleavage

provides a template for labeling the restriction fragments with biotin-14-dCTP, which will allow for the enrichment of Hi-C ligation products formed from crosslinked restriction fragments. The 5' overhang is filled in by the Klenow fragment of DNA polymerase I using equimolar amounts of all deoxyribonucleotides with the substitution of biotin-14-dCTP for dCTP. A small aliquot (10-20%) of the sample is not filled in but is otherwise treated the same as the Hi-C library to produce a standard 3C library. This 3C control library is used in subsequent quality control steps.

Both the Klenow fragment and the remaining HindIII enzymes are denatured with SDS while incubating at 65°C. The DNA fragments, which are still crosslinked to one another in chromatin complexes, are then ligated in a dilute reaction at 16°C. This dilute condition favors the intra-molecular ligation of fragments crosslinked within the same chromatin complex instead of ligation between fragments in different chromatin complexes. Since the ligation of these blunt ends is inefficient, the incubation time is set to 4 hours. The ligation of two completely filled in HindIII sites forms a NheI site: 5' – GCTAGC – 3'. This newly formed NheI site can be used to measure the efficiency of biotin fill-in by digestion of ligation junction amplicons with NheI.

DNA Purification.

The chromatin complexes containing the biotin-labeled ligation products are degraded by incubation with Proteinase K at 65°C. Proteins and lipids are removed from the ligation products with phenol pH 8.0:chloroform (1:1) in phase lock tubes. The DNA is then precipitated from the aqueous phase with sodium acetate pH 5.2 and 100% ethanol under centrifugal force. The precipitate may contain undesirable co-precipitates, such as salt and DTT, which can interfere with downstream molecular assays. To eliminate these co-precipitates, the DNA pellet is resuspended in 1X TE and washed with 1X TE using an Amicon 30 kDa molecular weight cutoff column. Contaminating RNA is degraded with RNase A treatment at 37°C. This procedure yields high-quality ligation products that are ready for downstream applications.

Quality Control of Hi-C Libraries.

The purified sample now consists of biotin-labeled ligation products formed between genomic restriction fragments that were adjacent to one another in physical space. Before proceeding, the library is assayed to ensure that it meets quality metrics. To begin with, the DNA is run on a 0.8% agarose gel to observe the distribution of fragment sizes in the library. The majority of the sample should run above 10 kb. Libraries that have a significant amount of degradation, indicated by a smear of lower molecular weight products, should be discarded because such libraries do not result in high quality data. Degradation

of Hi-C libraries can occur for several reasons, including: overly vigorous homogenization, not adding sufficient protease inhibitors during lysis, thermal degradation, or cell type specific effects, including endogenous nuclease activity. If degradation of libraries persists, a small aliquot can be removed prior to restriction enzyme digestion. DNA can be extracted and analyzed for degradation via gel electrophoresis. Library DNA is also quantified either by agarose gel electrophoresis or spectrophotometry. Usually, more than 100 µg of DNA should be recovered from 25 million mammalian cells.

To test the formation of ligation products, a standard 3C PCR reaction is performed to amplify a ligation product formed by two directly adjacent restriction fragments. Such a ligation product should be readily detected in a 3C or Hi-C ligation product library. The design of 3C primers is discussed elsewhere (Naumova et al. 2012). The resulting 3C amplicon is digested with HindIII and NheI in order to assess the proportion of the library that was filled in and labeled with biotin-14-dCTP (**fig. 2.2**). The complete 3C or Hi-C library of ligation products cannot be probed directly with NheI to assess biotin fill-in, because of

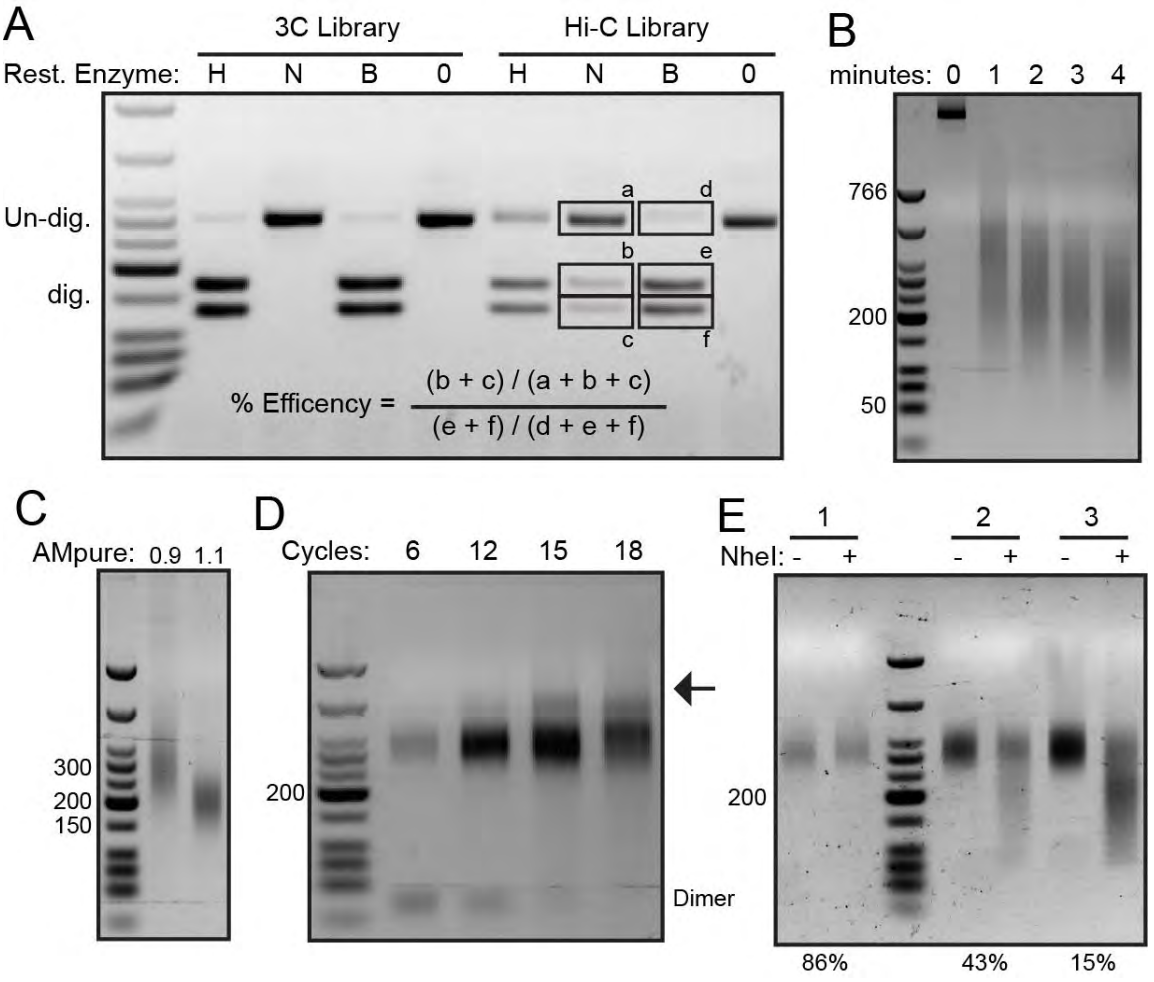


Figure 2.2. Relative ligation efficiency of Hi-C library. A) Digestion of a PCR amplicon generated by ligation of an adjacent pair of restriction fragments in both the 3C sample and the Hi-C sample. The amplicon was digested with HindIII (H), NheI (N), Both HindIII and NheI (B), or not digested (0). The amplicon when digested yields 2 products of different molecular weight. The molecular weight ladder is the Low Molecular Weight Ladder from NEB. **B)** The Hi-C library is fragmented using the Covaris 8700. A titration of fragmentation time is shown, starting with un-fragmented Hi-C library (lane 1) and increasing in the number of minutes of fragmentation (lanes 2-5). At 4 minutes (lane 5) the distribution of fragment sizes is ~50bp – 600bp. **C)** AMpure XP is used to fractionate the library. The 0.9x AMpure XP fraction includes molecules that are larger then ~150bp and the 1.1x fraction includes molecules that are between ~150bp – 300bp. **D)** The Illumina Paired-end graft sequences are added to the Illumina adapter modified Hi-C libraries using PCR with primers PE 1.0 and PE 2.0. These primers are partially homologous to the PE adapter, which was ligated to the Hi-C library. The gel shows a titration of the number of cycles. At higher numbers of PCR

cycles, higher molecular weight artifacts are produced (arrow). Sufficient amounts of DNA are produced at 12 cycles for this library. **E)** Three completed Hi-C libraries were digested with NheI. The shift of the size distribution of the library following digestion with NheI estimates the proportion of the library that consists of real Hi-C ligation products. A range of performances are shown, with library 1 showing poor performance, library 2 showing medium performance and library 3 showing good performance. The percentages below the each library are the percentages of dangling-ends that were tabulated after sequencing these libraries and mapping the reads to the genome.

the potential presence of endogenous NheI restriction sites within the ligated HindIII fragments. The amplicon generated from the 3C control library should digest completely with HindIII and not NheI, because the HindIII restriction site was not filled. In contrast, a significant fraction of the PCR amplicon obtained with the Hi-C library can be digested with NheI. Complete digestion of these PCR products is rarely achieved (**fig. 2.2a Lanes 4 and 8**). This is attributed mostly to point mutations made at the HindIII cleavage site during the PCR amplification. The use of high-fidelity DNA polymerases in the PCR reaction attenuates this problem. To quantify the fraction of molecules containing the NheI site, the digest is run on a gel and the bands are quantified with image analysis software. The ratio of NheI-digested products (**fig. 2.2a lane 7**) to the proportion of the tandem NheI and HindIII digested products (**fig. 2.2a lane 8**) is the proportion of the library that has been properly filled in with biotin. This efficiency of biotin incorporation can vary between libraries but is usually 20 – 30 percent. In cases where the filling-in efficiency is low (efficiency <5%), the library can be remade with an extended biotin fill-in incubation step and an overnight ligation with additional ligase. Increasing ligation time can drive the reaction to ligate more of those fragments that are biotinylated, yet this may also increase the frequency of intermolecular ligation of non-cross-linked molecules. The 3C control library is informative if few or no PCR products are formed from the Hi-C library. If little PCR product formed in the 3C control library as well, then the ligation, which is common between the 3C and Hi-C libraries, was inefficient.

However, if the 3C control library generates more PCR product than the Hi-C library, the biotin fill-in step was inefficient.

Biotin Removal From un-ligated Ends.

All subsequent steps pertain solely to the Hi-C library and not the 3C control library. Ligation of biotin-marked DNA ends is not as efficient as ligation of staggered ends. Thus, in a typical Hi-C experiment, a significant fraction of DNA ends do not become ligated. The biotin-labeled ends that have not been ligated together must be selectively removed so that they are not pulled down on the streptavidin-coated beads along with true ligation products. The strong 3' -> 5' exonuclease activity of T4 DNA Polymerase is used to remove nucleotides from the ends that have not been ligated together. A high enzyme concentration and low abundance of nucleotides drives the polymerase towards exonuclease activity over polymerization. dATP and dGTP are the only nucleotides added during this reaction to allow balancing nuclease and polymerase activity on the HindIII template strand after Biotin-dC removal which prevents complete degradation of the DNA. This reaction is stopped with the addition of EDTA. Then, the DNA is extracted with Phenol pH 8.0:Chloroform (1:1) and precipitated with sodium acetate pH 5.2 and 100% ethanol. The reaction is incubated on dry ice for 30 minutes and the precipitate is collected by centrifugation. The DNA pellets are then resuspended in water and the salt in the solution is washed out

three times with H_2O using a 30 kDa Amicon column. It is important to wash the Hi-C library thoroughly to remove salt because the sample will perform better in subsequent enzymatic reactions and during fragmentation and size fractionation.

DNA Fragmentation and Size Fractionation.

The rest of the protocol describes how to modify the biotinylated library of ligation products for sequencing with the HiSeq platform from Illumina. The ideal size of the library will depend on which sequencing platform will be used. A Hi-C ligation product size distribution of 150 – 300bp is optimal for most experiments because this size distribution works well for HiSeq cluster formation.

There are a variety of options available to fractionate DNA. The Covaris 8700 apparatus optimizes the amount of DNA in the relevant size range (**fig. 2.2b**) and is very reproducible. Regardless of which technology is used, the conditions for each system must be optimized to get the maximum amount of DNA as close to the preferred size for sequencing as possible.

To achieve an even tighter size distribution after fragmentation, AMpure XP mixture (Invitrogen) is used to fractionate the DNA based on size. This AMpure XP mixture includes Solid Phase Reversible Immobilization (SPRI) paramagnetic beads in a specific solution that precipitates DNA onto the beads. When AMpure XP is added to a sample, the DNA will precipitate onto the SPRI

beads, which can be recovered with a Magnetic Particle Separator (MPS). The relative proportion of the AMpure XP solution, after addition to the sample, will determine which DNA sizes precipitate onto the SPRI beads. Larger DNA molecules fall out of solution more easily than smaller ones, so less AMpure XP solution is needed to precipitate them. To purify the correct size fraction, low binding pipet tips are recommended and accuracy in pipetting is paramount to achieve precise proportions of AMpure XP solution relative to the sample volume.

Ligation products that are larger than 150bp are removed by the addition of 0.9x volumes of AMpure XP solution (**Fig. 2.2C lane 2**). The SPRI beads are harvested with the MPS. This higher molecular weight fraction should be kept in case another larger fraction is needed. To the remaining supernatant, more AMpure XP is added, bringing the relative total volume of AMpure XP solution to 1.1X of the original Hi-C sample. This second addition of AMpure XP is supplemented with additional SPRI beads which were harvested with the MPS to prevent saturation of beads with DNA and incomplete DNA capture. The SPRI beads in the 1.1x fraction bind DNA that is between 150bp and 300bp (**Fig. 2C Lane 3**) and the supernatant contains ligation products that are smaller then 100bp. The beads are harvested with the MPS and the supernatant is discarded. The SPRI beads with both the 1.1x fraction and the 0.9x fraction are washed with 70% ethanol to remove any remaining AMpure solution. After air-drying at 37°C, the DNA for both fractions is eluted with Elution Buffer (EB) from Qiagen. The

DNA is then analyzed and quantified on an agarose gel to ensure that the size distribution is as expected (**Fig. 2C Lane 3**).

End Repair and 'A' Tailing.

Shearing the Hi-C library causes asymmetric breaks in the DNA molecules and these broken ends must be repaired before Illumina adapters can be ligated. The Klenow fragment of DNA Polymerase I is used to fill in 5' overhangs while the T4 DNA Polymerase is used to both fill in 5' overhangs and to chew back 3' overhangs. Simultaneously, the T4 Polynucleotide Kinase is used to add a 5' phosphate to the Hi-C library to allow for ligation of Illumina sequencing adaptors. The Hi-C ligation products are purified from the reaction using Qiagen MinElute columns. Klenow (exo-) is then used to adenylate the 3' end of the fragment, which will allow for the ligation of the Illumina PE adaptors.

Streptavidin Pull-down of Biotinylated Hi-C Ligation products.

In order to minimize non-specific pull down of DNA during the streptavidin pull-down of Biotinylated Hi-C ligation products, all steps are done in low binding tubes and with low binding pipette tips. The biotinylated Hi-C ligation products are mixed with My-One C1 streptavidin bead solution (Dynabeads). The volume of Dynabeads added should relate to the amount of DNA in the sample. Using a

volume of Dynabead solution in the range of 2 – 5 uL per 1 ug of total DNA (as quantified at the end of the size fractionation step 1.7) should provide an excess of beads relative to DNA. Once the DNA is bound to the beads, they are washed to remove non-specifically binding DNA and the buffer is exchanged with 1X T4 Ligation Buffer from Invitrogen to prepare for ligation of the Illumina adaptors.

Paired-End Adapter Ligation and Library Amplification.

Ligation of the Illumina Paired-end Adapters is done while the Hi-C library is bound to the streptavidin beads. The adsorption of the DNA to the beads increases the efficiency of adapter ligation by decreasing the mobility of the DNA fragments and facilitates removal of un-ligated oligonucleotides. After ligation of the adapters, the sample is washed to remove the un-ligated oligonucleotides and to exchange the buffer. The beads are finally resuspended in 20 uL NEBuffer 2.

It is important to PCR amplify the library with as few cycles as possible (9-15 cycles). This ensures linear amplification without creating PCR artifacts (**Fig. 2.2D arrow**), while producing enough product to successfully sequence the library (**Fig. 2.2D lane 3**). It is preferable to pool multiple PCR reactions rather than to increase the number of PCR cycles. A test PCR should be run to titrate the number of cycles needed (**Fig. 2.2D lanes 2-5**). If linear amplification cannot be achieved in the target range of cycles, the volume of beads for each reaction

can be adjusted. If the amount of DNA required for sequencing cannot be obtained with ~18 cycles or fewer, it will be necessary to repeat the experiment starting with a higher number of cells.

After the optimal number of cycles has been determined, multiple PCR reactions are performed to produce the amount of library needed for sequencing (about 50 ng of DNA). Usually, 5 reactions will produce sufficient amounts of DNA. The amplified DNA is then pooled and concentrated, and the primer dimers are removed by using 1.8x AMpure XP as before (step 2.7). The purified DNA can be quantified either with agarose gel electrophoresis or spectrophotometrically. It is advised to verify the size and to accurately quantify the Hi-C library with a Bioanalyzer (Agilent Technologies; Santa Clara, CA) prior to sequencing.

Final Quality Control and Library Quantification.

The efficiency of the removal of biotin in un-ligated restriction fragments in step 2.6 varies between samples. This is attributed to variation in ligation efficiency and also may relate to the incorporation of non-junction biotins at small nicks in the DNA sequence. To gauge the fraction of true ligation products in the pulled-down DNA, a small portion of the final amplified library is digested with NheI (**Fig. 2.2E**). This digestion cleaves the library at NheI sites that resulted from the ligation of biotin-labeled restriction fragments. Upon digestion of the

library with NheI there should be a shift in the distribution of fragment sizes for the library. The degree to which the distribution shifts is inversely proportional to the number of biotinylated but un-ligated “dangling” ends in the sample. By quantifying the amount of sample that shifted to a lower molecular weight, the proportion of dangling ends can be approximated. Most libraries typically consist of approximately 10-45% of dangling ends. Libraries that have more than 80% dangling ends will most likely not be profitable to sequence.

It is important to note that this approach provides only a rough approximation of the proportion of dangling ends. Endogenous NheI sites present within 500 bp of a genomic HindIII site may cause an overestimation of the fraction of digested products. However, there are also true ligation products that cannot be cut, because of fill-in errors. Further, the NheI site can be located near one end of the sheared DNA molecule. When digested, such molecules will not shift appreciably in molecular weight, leading to an underestimation of the proportion of dangling ends in the Hi-C library. Although these factors may confound the interpretation of the results, in general we have found that this estimate of dangling end percentage is a reliable indicator of the quality of the library (**compare fig. 2.2E sample 1 to fig. 2.2E sample 3**).

2. Expected Results and Discussion

Sequencing of Hi-C libraries.

In principle, the final Hi-C library of paired fragments can be sequenced using any platform that will allow both ligated sequences to be mapped to the genome, either by long reads that will read through the NheI junction (Roche 454) or by paired-end or mate-paired reads (Illumina GA and HiSeq platforms and Life Technologies SOLiD). To check the quality of a library before high-throughput sequencing, a small subset of library molecules can be cloned and sequenced using traditional Sanger sequencing, where the resulting long sequencing read will pass through the ligation junction, allowing identification of interacting pairs of loci. We have found that Illumina paired-end sequencing with 36 or 50 bp reads is an effective way to identify a large number of interacting fragment pairs. Longer read lengths (75 bp or 100 bp) may improve mappability for repetitive regions. However, with an average library size of 250 bp after sonication and size selection, such longer reads are likely to pass through the ligation junction into the partner fragment. Unless reads are truncated at such ligation junctions (5'-AAGCTAGCTT-3' in the case of HindIII), these longer reads may be unmappable to the (linear) reference genome, where the ligated fragments are not neighboring. Therefore, we find that 50 bp paired-end reads are optimal for Hi-C library sequencing.

Sequence Read Mapping and Filtering.

The sequenced Hi-C reads can be mapped to the genome with any short read sequence alignment algorithm (Bowtie, Maq, Eland, Novoalign, etc.). Analyzing the position and direction of sequenced reads relative to restriction sites provides information about the types of molecules present in the Hi-C library and the overall quality of the library (**Fig. 2.3A**). If the two reads occur in the same fragment, they either represent a self-circularized ligation product or an unligated dangling end product. Typically, self-circles comprise 0.5-5% of the molecules in a Hi-C library. The proportion of self-circles increases with decreased crosslinking because there is less competition for ligation of the two ends of that fragment by other fragments. Thus, changing the formaldehyde concentration or crosslinking time may change the proportion of self-circles in the final library. Dangling ends may comprise 10-45% of a successful Hi-C library. The T4 exonuclease activity during the Biotin Removal step of the protocol (1.6) is essential for minimizing the proportion of dangling ends.

In contrast to the dangling ends and self-circles, valid interaction pairs map to different restriction fragments and face toward the restriction site. If the paired ends of multiple molecules have the same size and map to the same genome position at both ends, only one copy of this interaction should be considered. Such redundant molecules are unlikely to occur by chance and thus may result from PCR over-amplification. If more than 5% of the library of valid pairs has redundant molecules, concern is raised that the library may have been amplified by too many PCR cycles. A high percentage of redundant molecules

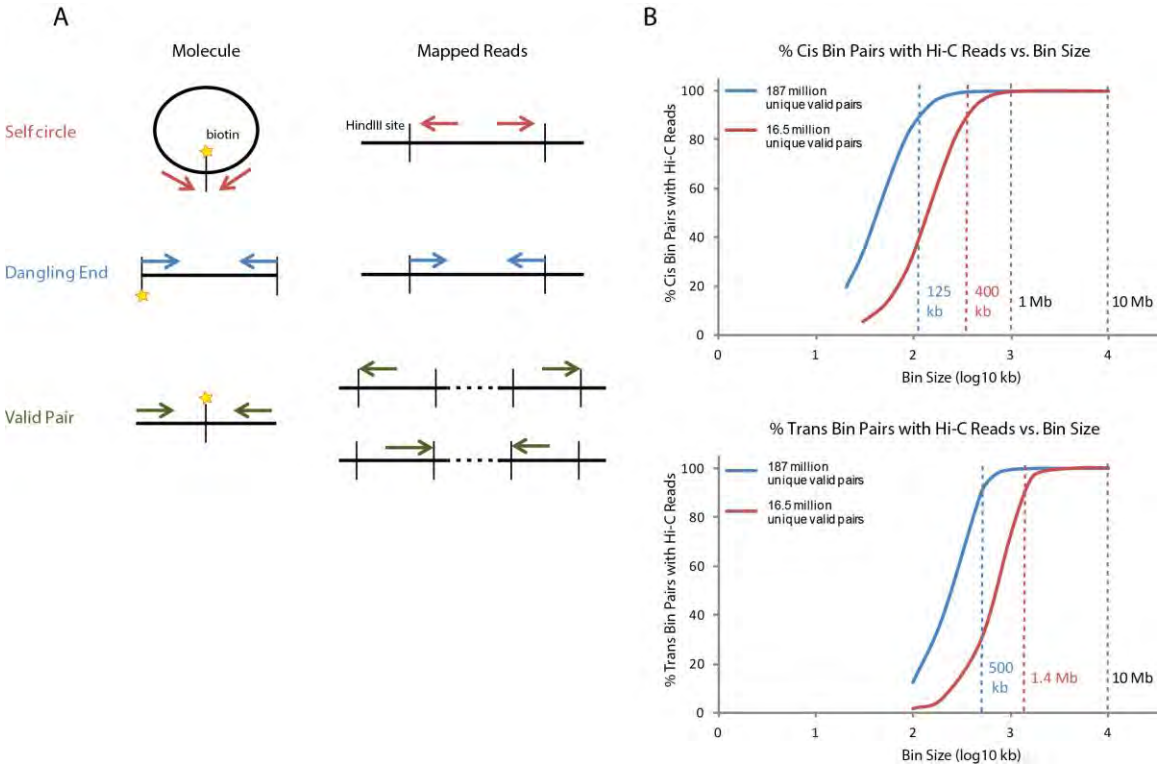


Fig 2.3. Hi-C sequence mapping and binning considerations. A) Different types of molecules in the Hi-C library (left) lead to different orientations of mapped reads relative to restriction sites (right). Mapped reads (colored arrows) facing outward in the same fragment come from self-circles (top); Reads facing inward in the same fragment arise from dangling ends (middle); Reads from different restriction fragments and facing toward a restriction site arise from valid interaction pairs (bottom). **B)** Relationship between sequencing depth and choice of bin size. Each graph shows the percentage of *cis* (top) or *trans* (bottom) bin pairs that contain at least one mapped read from a valid interaction pair (y-axis) for each different bin size (x-axis). Colored dotted lines indicate the bin size at which 90% of bin pairs contain at least one valid pair read for a Hi-C library with a high (blue) or low (red) number of total unique valid pairs after sequencing.

may reflect a lack of complexity in the Hi-C library that may arise from the use of an insufficient number of cells in the Hi-C experiment. With a small number of cells, the total number of captured interactions may be small. Thus, some interactions are over-amplified by PCR (and thus appear as redundant molecules) while others are never observed, resulting in an interaction map composed of a very small number of high-copy interactions and a very large percentage of never-observed interactions.

Data Resolution, Binning, and Normalization.

Unique valid interaction pairs (non-redundant, true ligation products) can now be used as a measure of the frequency of physical contact between each pair of loci in the genome. Assuming that every restriction fragment could ligate to any other, there are on the order of 10^{11} possible HindIII restriction fragment pairs in the human genome. Thus, it is difficult to generate a Hi-C library with enough complexity or sequence depth to cover all possible restriction fragment interactions. Additionally, many reads from a typical Hi-C experiment represent proximal interactions (10-15% of reads represent interactions of genomic regions separated by less than a megabase). Thus, thoroughly sampling distant interactions requires a relatively large increase of sequencing depth (**Fig. 2.4B and 2.4C**) (Dixon et al. 2012b). In order to gain statistical power, it is useful to pool numbers of reads within larger genomic regions (bins), before

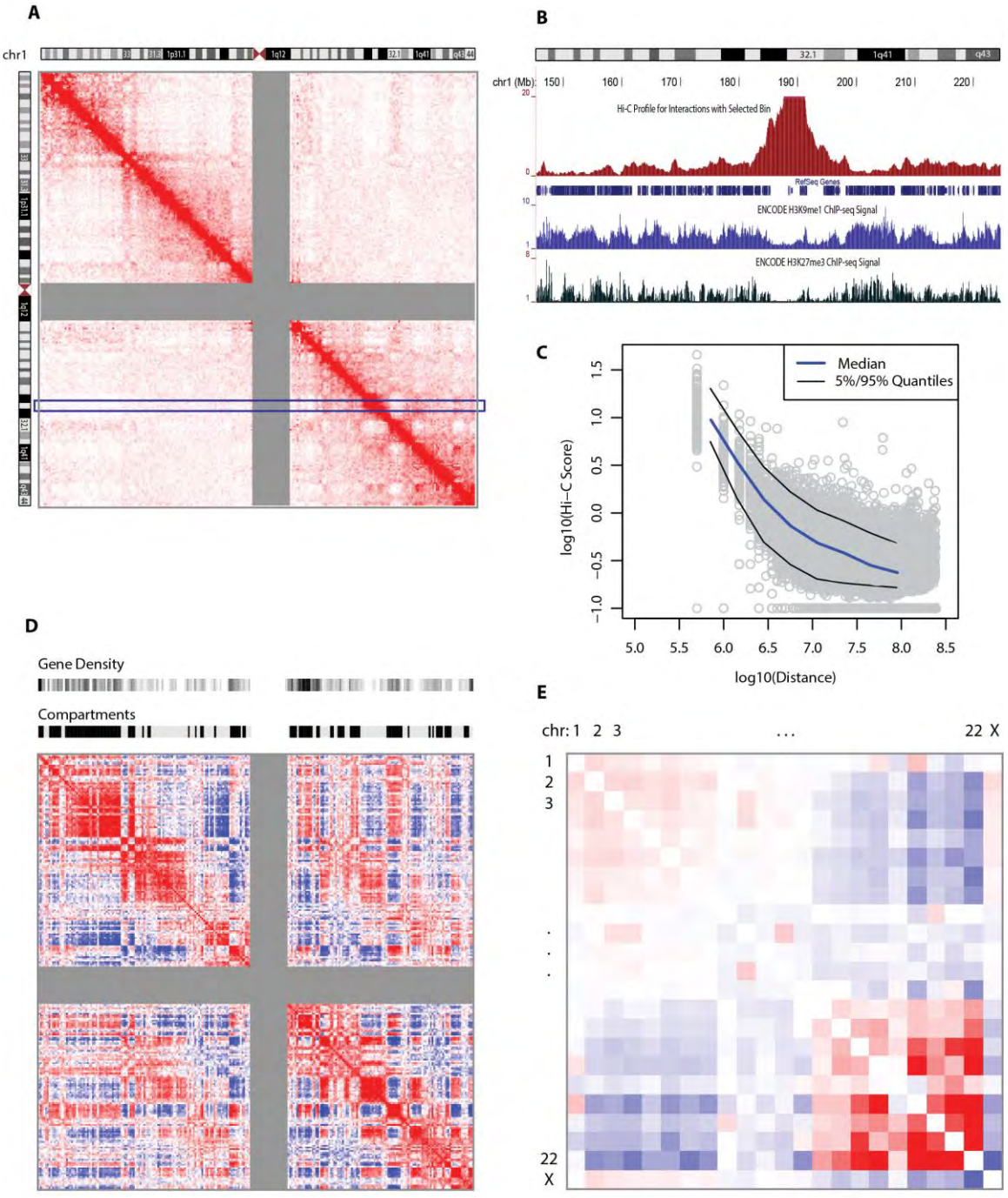


Fig 2.4. Hi-C data visualization and analysis. A) A heatmap of interactions between all 1 Mb bins along chr1 for GM06990 cells. The intensity of red color corresponds to the number of Hi-C interactions. **B)** A "4C profile" derived from one row of the Hi-C heatmap (blue box in A) showing all interactions between a

fixed 1 Mb location at 190 Mb on chr1 and the right arm of chr1. H3K9me1 and H3K27me3 tracks are displayed below as examples of other genomic datasets that can be compared with such an interaction profile. **C)** The log10 of the Hi-C interaction counts of each pair of bins along chr1 is plotted versus the log of the genomic distance between each pair of bins. The median value of datapoints in the graph is indicated by a blue line while the 5% and 95% quantiles are shown as thin black lines. The slope of the median line from 500 kb to 10 Mb is -1, following the relationship expected for a fractal globule polymer structure of the chromatin. **D)** Red and blue "plaid" patterns show the compartmentalization of chr1 in two types of chromosomal domains. The data from A were transformed by first finding the observed interactions over the expected average pattern of decay away from the diagonal and then calculating a Pearson correlation coefficient between each pair of rows and columns. Regions highly correlated with one another in interaction are colored red and are likely to be classified by principle components analysis into the same compartment as shown above (black bands = open chromatin compartment; light grey bands = closed chromatin compartment). The compartment assignments correlate with the gene density profile, shown above the compartment profile (high gene density = black; low gene density = white). **E)** Whole chromosome interaction patterns show that longer chromosomes (chr1-10, chrX) are more likely to interact with one another and not with shorter chromosomes (chr14-22). The observed number of interactions between any pair of chromosomes is divided by the expected number of interactions between those chromosomes given the total number of reads for either chromosome in the whole experiment. Red indicates an enrichment of interaction as compared to expected values while blue indicates a depletion of interactions between two chromosomes.

further analyzing the data. Larger bins will contain more reads and thus have more discriminatory power, but at the cost of lowering the resolution of the data. The optimal bin size, and therefore the resolution at which the interaction data can be analyzed, depends on the sequencing depth and the linear separation of the genomic regions under consideration (**Fig. 2.3B**). Since intra-chromosomal ("*cis*") interactions, are more likely to occur and thus more frequently observed, a smaller bin size may be effective when analyzing interactions within one chromosome. On the other hand, a larger bin size may be necessary to distinguish the less frequent inter-chromosomal ("*trans*") interactions. For mammalian genomes, we find that at least 7 million unique valid pairs are necessary to analyze higher order features like the positioning of whole chromosomes in the nucleus, polymer scaling analyses, or the compartmentalization of genomic regions into open and closed chromatin (see section 3.4). To look at interactions in *cis* between particular genomic bins at a scale of 100 kb, we recommend obtaining more than 100 million unique valid pairs. On the other hand, for a library with 100 million reads total, the minimum bin size appropriate for analyzing *trans* data would still be about 1 Mb.

The raw counts of Hi-C reads in different genomic bins will be influenced by several experimental biases, such as the mappability of the sequence in the bin, the number and length of restriction fragments in the bin, the propensity of the DNA in the bin to be PCR-amplified and sequenced, etc. (Yaffe and Tanay 2011b). For example, in the raw data, it may appear that a certain region is

depleted for interactions, but in fact there are simply fewer restriction fragments in the region that could be forming interactions. Therefore, before proceeding with additional analyses, it is important to correct the Hi-C interaction map to account for these possible biases. Several approaches have been developed to correct Hi-C data. A method developed by Yaffe and Tanay explicitly defines each bias separately (e.g. fragment length) and then attempts to correct for these biases (Yaffe and Tanay 2011a). Alternatively, we have developed an approach that takes advantage of the genome-wide nature of the Hi-C interaction data directly to estimate detection bias without the need to know all factors that could bias the analysis, or the need to explicitly define their impact (Imakaev et al. 2012). Our coverage correction uses a mathematical approach that normalizes the "visibility" of each genomic bin, defined as the total sum of interaction reads detected between a single bin and the rest of the genome (the sum of a column or row in the matrix of (**Fig. 2.4A**). This approach assumes that, while different genomic regions may have different distributions of interactions with other proximal or distal sequences across the genome, each restriction fragment will form some interaction, whether it be only with neighboring fragments or with many distal regions. Thus, in the absence of experimentally induced biases, the total sum of interactions across the genome of a given genomic bin should be equal. We have found that our coverage correction produces results equivalent to the Yaffe and Tanay correction, with less computation time (Yaffe and Tanay 2011a; Imakaev et al. 2012). Bias correction procedures such as these will also

account for the number of reads in the dataset, making it possible to compare such normalized Hi-C scores across experiments.

Visualization of Hi-C Data and Basic Expected Results.

Once the data have been binned and normalized, genome spatial organization can be visualized and analyzed in a variety of ways, according to the goals of each particular study. To observe the patterns of all pairwise interactions, a heatmap matrix of normalized interaction values can be constructed (**Fig. 2.4A**). In any successful Hi-C experiment, an interaction heatmap should show a strong diagonal of interactions between proximal genomic regions and an overall exponential decay in interaction signal over increasing distance. The thickness, or width, of the diagonal relates to the level of chromatin compaction along the chromosome.

Each row or column of this heatmap (bin) details the interactions between one genomic region and the rest of the chromosome or genome (effectively a "4C profile"). These interaction profiles of single loci can be directly compared with other genome-wide datasets to investigate, for example, the relationship between 3D conformation and epigenetic features or gene regulation. If the Hi-C data is of low-resolution (~1 Mb bins), this 1D profile is suitable for comparisons with genomic features that occur over broad genomic regions, such as certain histone modifications (**Fig. 2.4B**). Profiles extracted from a higher resolution

dataset (< 100 kb bins) could be compared with finer-scale features, such as transcription factor binding patterns. As a specific example, 1D Hi-C profiles have been used to investigate the relationship between the 3D interactions and translocations for a particular genomic locus (Zhang et al. 2012b).

Hi-C data can also be used to infer the polymer structure of chromatin. Different DNA polymer structures (fractal globule, equilibrium globule, etc.) have different predicted relationships between contact probability and linear genomic distance (Lieberman-Aiden et al. 2009b; Mirny 2011), and such relationships can be determined by comparing the number of Hi-C interactions between all pairs of regions with the genomic distance between each pair (**Fig. 2.4C**).

Additionally, Hi-C data can reveal the “compartmentalization” of the genome into regions of open and closed (active and inactive) chromatin, as described previously (Lieberman-Aiden et al. 2009b). Hi-C interactions tend to be much stronger within one such compartment than between compartments. Classifying all genomic regions into compartments summarizes the data as neighborhoods of interaction that can be visualized and compared to other genomic datasets (**Fig. 2.4D**). By identifying changes in compartment identity after a certain perturbation, Hi-C data can be used to investigate which regions of the genome are rearranged spatially in response to a treatment, pathology, or developmental transition.

Lastly, the relative positions of whole chromosomes in the nucleus can be visualized on a larger scale by comparing the observed number of interactions between each pair of chromosomes to the number of interactions that would be expected if reads from each trans interaction were spread evenly across all chromosomes based on their respective sizes (**Fig. 2.4E**; specific calculations are detailed in (Zhang et al. 2012b)). These results can be compared to the organization of chromosome territories as observed by imaging methods (Cremer and Cremer 2010).

Comparison to Other Methods.

Once a Hi-C experiment has revealed salient patterns of genome organization, interesting findings can be validated and further investigated using other approaches. As described earlier, imaging spatial relationships between regions of the genome using FISH is not suited to the study of genome wide chromatin structure, but it is a powerful tool for visualizing the relationships between specific genomic loci (Solovei et al. 2002). A FISH experiment with probes specific to given loci can reveal the 3D distance between loci on a per cell basis. The average 3D distance between loci measured is typically inversely related to the Hi-C signal (Lieberman-Aiden et al. 2009b).

If a Hi-C experiment suggests that a certain region of the genome (on the scale of 10 Mb) undergoes particularly interesting changes in conformation

between conditions of interest, such a region can be explored at higher resolution using a 5C experiment (Dostie et al. 2006b; Dostie and Dekker 2007; Ferraiuolo et al. 2012a) . 5C probes can be designed for every possible interaction in the region identified by Hi-C and then the 5C library, generated from a 3C library in a manner similar to Hi-C, can reveal further high-resolution (typically at a resolution of single restriction fragments) details of the spatial organization of this region.

Conclusion.

The highly parallel nature of the Hi-C method produces genome-wide interaction maps. As such, Hi-C provides a unique and powerful tool to study nuclear organization and, chromosome architecture. Thus, Hi-C data will add a spatial context to biological inquiries and will facilitate the discovery of the fundamental principles of gene regulation, nuclear partitioning, and the biophysics of chromatin dynamics. Hi-C does not capture the fine detail of sub-nuclear compartments like electron microscopy nor can it measure the dynamics of interactions between multiple genomic loci like fluorescent microscopy. It does, however provide the ultimate connectivity between the genomic sequence and spatial conformation. The genome-wide power and versatility of Hi-C makes it ideal for the study of the basic biology of genome organization and its implications for health and disease.

CHAPTER III

Measuring Chromatin Structure in Budding Yeast

Contributions

The text and figures for this chapter are from: Belton, J.-M., Dekker, J. "Measuring Chromatin Structure in Budding Yeast." Budding Yeast: A Laboratory Manual. Cold Spring Harbor Laboratory Press. – *Accepted for Publication.* The authors are listed below: Jon-Matthew Belton, Job Dekker

Abstract

Chromosome Conformation Capture (3C) based techniques have changed the way the conformation of chromatin and its relationship to other molecular functions of the cell can be studied. These techniques are used to determine the spatial arrangement of chromosomes in organisms ranging from bacteria to human. Application of 3C-based method to the analysis of chromosome organization in model organisms with small genomes and for which powerful genetic tools exist, such as budding yeast, allows analysis of mechanisms of chromosome folding at very high resolution and at relatively low cost. Different variations of the basic 3C approach (e.g. 3C-PCR, 5C, and Hi-C) can be used depending on the scope and goals of a given experiment. Here we describe detailed protocols for performing these methods in yeast. In addition the power of yeast genetics allows the opportunity to elucidate mechanisms which establish or maintain chromatin structure. Study of these mechanisms at high resolution in a genetically perturbable system will further our understanding of these fundamental processes in higher eukaryotes as well.

Introduction

Ever since chromosomes were first described, understanding the structures of these enormous molecules and how these structures are relevant to the many functions of the genome, i.e. gene expression, DNA replication, and chromosome transmission, have been fundamental questions in biology. For decades light and electron microscopy have been instrumental in these studies. In 2002 a molecular approach was developed to complement the imaging-based methods. This method is Chromosome Conformation Capture (3C) (Dekker et al. 2002b), which relies on formaldehyde crosslinking to detect the physical association of genomic loci in three-dimensional (3D) space (**Fig. 3.1a**). Briefly, in 3C chromatin is crosslinked with formaldehyde followed by digestion of the chromatin with a restriction enzyme. The restriction fragments that are crosslinked to one another are then ligated together. These chimeric DNA molecules are then purified, and the result is a library of DNA molecules which represent 3D spatial interactions between those restriction fragments. This library can then be analyzed in a variety of ways including PCR and deep sequencing. Counting how often a pair of restriction fragments is found crosslinked and then ligated to one another, indicates the likelihood that those two fragments of chromatin interact in the cell. PCR primers designed to chromatin fragments of interest can be used to detect and quantify specific interactions one at the time (**Fig. 3.1b**). High throughput versions of 3C such as 3C-on-ChIP/Circular 3C (Simonis et al. 2006b; Zhao et al. 2006b), Chromosome

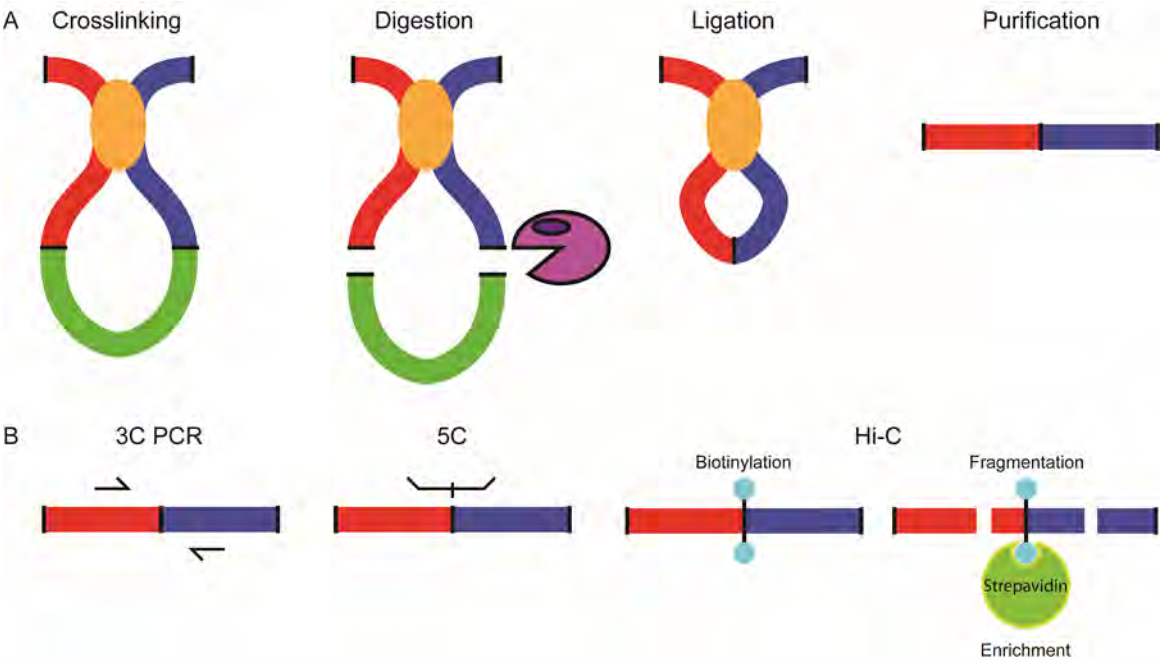


Figure 3.1. Schematic outlines of the 3C, 5C and Hi-C techniques described in this chapter. (a) This panel shows the steps that are common to all 3C based techniques. First the cells are crosslinked with formaldehyde. This process covalently links protein-DNA complexes in the cell nucleus together. The DNA is then digested with a restriction enzyme of choice. This breaks up the crosslinked chromatin and frees covalently linked protein-DNA complexes into solution. These complexes are diluted in a ligation reaction. The dilution favors ligation between two chromatin fragments that are attached to the same complex. After ligation the DNA is purified. (b)There are different ways to detect these 3C ligation products. Locus-specific PCR primers can be designed to detect specific interactions of interest. To detect many interactions at once, 5C uses pools of hundreds up to thousands of short oligonucleotides which hybridize to ligation products of interest and are then ligated together. The “carbon-copies” are analyzed by high through-put sequencing to quantify the interaction frequencies. For unbiased and genome-wide analysis using Hi-C the chromatin fragment ends are first biotinylated after digestion but before ligation. This marks the ligation junctions. After ligation DNA is fragmented by sonication and the molecules that contain a biotin at the ligation junction are enriched by pulling down with streptavidin beads. The resulting short molecules are then analyzed directly by high throughput DNA sequencing.

Conformation Capture Carbon Copy (5C), and Hi-C have subsequently been described to detect chromatin interactions for single loci genome-wide (4C), across targeted regions of the genome (5C) (Dostie et al. 2006a) or whole genomes in an unbiased manner (Hi-C) (Lieberman-Aiden et al. 2009a). 4C employs inverse PCR to detect all loci located to a single locus of interest. 5C uses short oligonucleotide probes to enrich a 3C library for interactions of interest, and Hi-C is a whole genome conformation capture assay which utilizes biotin incorporation to enrich for ligation junctions (**Fig. 3.1b**). These techniques have revolutionized the way we study and think about the structure of chromosomes.

Studies of chromatin conformation, using 3C-based techniques, have revealed many aspects of chromatin behavior. It is clear that chromatin is not randomly dispersed in the eukaryotic nucleus. In higher eukaryotes (e.g. mouse and human) the genome is organized hierarchically. The highest level of organization is that of chromosome territories which are on the order of 10's - 100's of Mb in size. Chromosome territories are formed because the chromosomes only rarely intermingle with other chromosomes. The next level of hierarchy of chromatin is compartments. These compartments are formed by the association of large (1's -10's Mb) regions of chromatin that have similar properties (active and inactive) and are spatially separated from each other in the genome (Lieberman-Aiden et al. 2009a; Zhang et al. 2012a). Compartments are further subdivided into Topologically Associating Domains (TAD's) which are

regions (100's kb – 1'sMb) where chromatin is preferentially interacting with itself but not with neighboring regions (Dixon et al. 2012a; Nora et al. 2012). These hierarchical levels of chromatin structure play a role in many biological functions such as gene expression (Bau et al. 2011; Sanyal et al. 2012; Jin et al. 2013; Symmons et al. 2014) and DNA repair (Zhang et al. 2012a). Defects in this organization have also been implicated in human pathology, e.g. in the premature aging disease Hutchinson-Gilford Progeria Syndrome (McCord et al. 2013).

The properties of chromatin at finer scales, e.g. at the sub-TAD scale in mammals, are less well characterized, in part because it is costly to perform comprehensive chromatin interaction studies at the resolution of single restriction fragments in organisms with large genomes. Since the number of possible pairwise chromatin interactions scales with the square of the number of restriction fragments in a genome, organisms with small genomes present a unique opportunity to study the conformation of chromatin at high resolution in a cost-effective manner. Budding yeast, *Saccharomyces cerevisiae*, has a genome size of ~12Mb distributed over 16 chromosomes, is genetically tractable and its genome can easily be manipulated. This, coupled with the vast depth of knowledge know about the physiology of buddying yeast, allows for the analysis of chromatin interactions in contexts that are not readily available in higher eukaryotes, including an array of unique environmental conditions. Further, the roles of specific proteins suspected to be involved in chromatin organization can

be straightforwardly studied. In addition, yeast cultures are easily synchronized to study changes in chromosome conformation during the cell cycle.

The type of 3C-based technique that should be employed for a given study depends on the question and scope of the experiment. All of these techniques rely on formaldehyde crosslinking of chromatin and subsequent proximity ligation of crosslinked restriction fragments in order to detect chromatin interactions. The only difference is the manner in which the products are detected.

3C-PCR is applicable when there are only a few (key) regions/loci of interest for which one would like to study their physical associations. Typically, experiments with less than 50 restriction fragments to be analyzed can be readily carried out using classical 3C followed by PCR. 4C is suitable for studies in which one is interested to see what a single locus of interest, e.g. a centromere, interacts with genome-wide. For more comprehensive studies of target regions, e.g. whole chromosomes, but that are not genome-wide, 5C is more applicable. 5C allows for the high-throughput detection of chromatin interactions in a single reaction (Dostie et al. 2006a). Short probes are used to hybridize to ligation junctions of interest in a 3C library. When two chromatin fragments of interest are ligated to each other in a 3C library the probes anneal adjacent to one another and are then ligated together with Taq Ligase. This, in effect, makes a "carbon-copy" of the ligation product. These products are then detected with

high throughput sequencing of the short ligated probes. The types of questions that can be addressed with 5C depend on how the probes are designed. The probes can be arranged for all restriction fragments along a chromosome which would yield comprehensive information of the spatial distribution of the entire chromosome. This information could be used to generate three-dimensional (3D) models of the chromosome (Bau and Marti-Renom 2011; Bau et al. 2011). The probes could also be designed so that one set hybridizes to restriction fragments containing regulatory elements and another set hybridizes to fragment overlapping genes. This design would allow for a comprehensive analysis of the spatial interactions between regulatory elements and their genes (Sanyal et al. 2012). Many other variations of these two generic probe designs are possible. 5C experiments typically use 50-5000 probes to detect hundreds to millions of chromatin interactions in parallel.

When the goal of the experiment is to obtain information regarding the spatial organization of a complete genome, then Hi-C is the technique of choice. In Hi-C, the ends left after restriction digestion are filled in with biotinylated nucleotides prior to ligation. As a result ligation junctions are labeled with biotin and can be purified in an unbiased manner using streptavidin-coated agarose beads. Purified ligation junctions are then analyzed and quantified using high throughput sequencing to obtain an unbiased genome-wide interaction map.

Chromosome Conformation Capture (3C) in Budding Yeast

3C takes advantage of formaldehyde crosslinking to measure the spatial association of two pieces of chromatin (**Fig. 3.2a**). The method begins by whole-cell formaldehyde fixation of the chromatin, followed by grinding the cells in liquid nitrogen to lyse them. Fixation of whole cells and nitrogen grinding is preferable to fixation of spheroplasts because spheroplasting process may affect cell integrity and nuclear organization. Next the chromatin is solubilized with SDS and incubated at 37˚C. This opens the chromatin and makes it accessible for digestion. A type II restriction endonuclease is used to fragment the chromatin. Cross-linked DNA fragments are then ligated together in dilute conditions which helps prevent inter-molecular ligation. The crosslinks are reversed by degrading the proteins with Proteinase K and incubating the reaction at 65˚C overnight. The remaining DNA is then purified by standard phenol:chloroform extractions followed by ethanol precipitation.

The resulting molecules are chimeric DNA molecules composed of restriction fragments which can be separated by large genomic distances, or that can be located on different chromosomes, but that were close enough in 3D space to be crosslinked. PCR primers are used to detect chromatin interactions of interest. The 3C PCR primers are designed 100-200bp upstream of the restriction cut site on the fragments of interest. All 3C primers are designed to hybridize to the "-" strand (**Fig. 3.2b**). This ensures that amplification only occurs

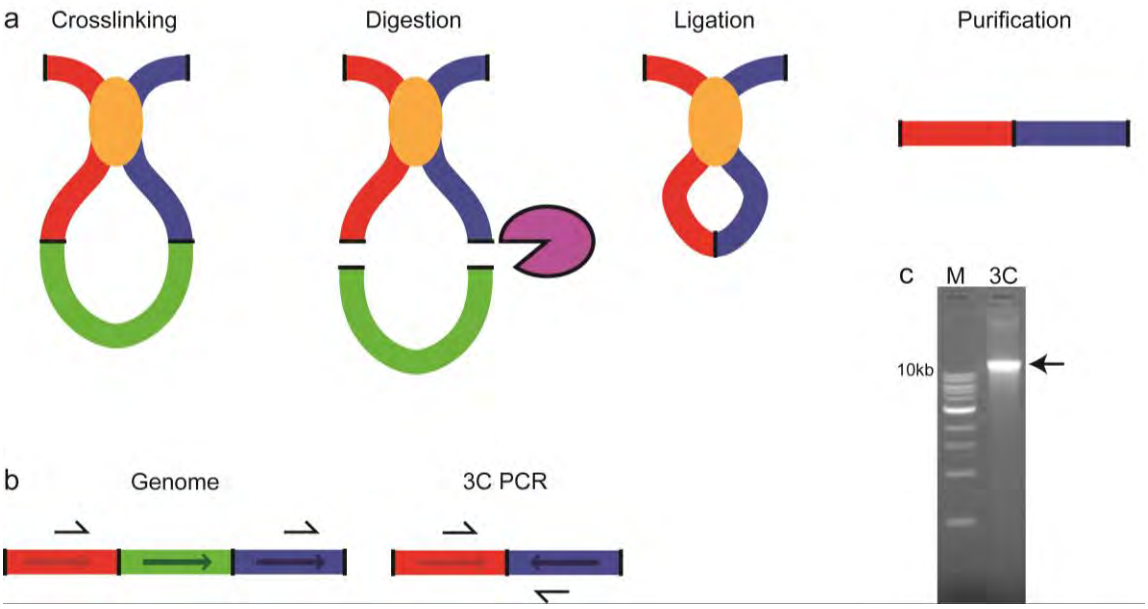


Figure 3.2. Overview of the 3C protocol, 3C primer design, and 3C quality control. (a) A 3C library is made by crosslinking chromatin with formaldehyde, digesting the chromatin with a restriction enzyme of choice, ligating crosslinked chromatin fragments together, and purifying the chimeric ligation products which represent spatial interactions between pairs of restriction fragments. (b) 3C primers are designed 100-200bp upstream of genomic restriction fragments. Each primer is designed to hybridize to the "-" strand so that PCR amplicons will only be formed when the two fragments of interest have been digested. This ensures that there is only detection of digested ligation products. (c) The 3C library run on a 0.8% agarose gel runs at ~12kb, indicated by the arrow. "M" is a 1Kb DNA ladder (New England Biolabs, N3232S).

when the fragments have been digested and then ligated in the opposite (head-to-head) orientation. This design eliminates false positive signal that can form by circularization of partial digestion products. Further discussion on 3C primer design and experimental set up can be found in (Dekker 2006; Naumova et al. 2012).

Materials

Reagents

1X Restriction Enzyme Buffer (which matches the chosen restriction enzyme)

1X TE

10X Ligation Buffer

Agarose

Appropriate growth media for the yeast strain of interest

ATP (100 mM)

Bovine Serum Albumin (BSA)(10 mg/ml)

Chloroform (100%)

Dry ice

Ethanol (100%)

Formaldehyde (37%, Fisher BP531-500)

Glycine (2.5M, filter sterilized)

Liquid Nitrogen

Phenol (pH 8.0):Chloroform (1:1)

Protienase K (10 mg/ml in 1X TE)

Restriction Enzyme (20,000 units/ml)

RNase A (10mg/ml)

Sodium Acetate (3M, pH 5.2)

Sodium Dodecyl Sulfate (SDS, 1% w/v)

Sodium Dodecyl Sulfate(SDS, 10% w/v)

Sterile Water

T4 DNA Ligase (1U/µl Life Technologies, 15224-017)

Triton X-100 (10% v/v)

Equipment

15 ml Amicon 30KDa column

15 ml conical tubes

15 ml Phase Lock Light tubes (5prime, 2302840)

16˚C water bath or thermomixer

1.7 ml microcentrifuge tubes

200 ml Conical Centrifuge tubes

250 ml Screw cap tubes for high speed spinning.

35 ml Screw cap tubes for high speed spinning.

37˚C water bath or thermomixer

50 ml conical tubes

50 ml Phase Lock Light tubes (5prime, 2302860)

65˚C water bath or thermomixer

96 well PCR plate

Glass slides and cover slips for microscopy

High Speed Centrifuge (up to 18,000xg)

Large centrifuge capable of spinning 18,000x *g*

Microscope with 100x objective

Mortar and pestle

Spectrometer to measure the Optical Density at 600nm

Table top centrifuge capable of spinning 3100x *g*

Thermocycler

Variable temperature incubator for growing yeast

Variable temperature shaking incubator for growing yeast

Vortex

Recipes

1X TE:

Stock Solution	Amount for 1L	Final Concentration
1 M Tris-HCl pH 8.0	10 ml	10 mM Tris-HCl pH 8.0
0.5 M EDTA pH 8.0	2 ml	1mM EDTA pH 8.0
De-ionized water	To 1L	

10X Ligation Buffer: *Store at -20˚C in 15 ml aliquots.*

Stock Solution	Amount for 1L	Final Concentration
1 M Tris-HCl pH 7.5	500 ml	500 mM Tris-HCl pH 7.5
1 M $MgCl_2$	100 ml	100 mM $MgCl_2$
1 M Dithiothreitol	100 ml	100 mM Dithiothreitol

De-ionized water	To 1L	

3M Sodium Acetate pH5.2 (1L): Dissolve 246.1g of Sodium Acetate in 500 ml of de-ionized water. Adjust the pH to 5.2 with Glacial Acetic Acid. Allow the solution to cool overnight. Adjust the pH once more to 5.2 with Glacial Acetic Acid. Adjust the final volume to 1L with de-ionized water and filter sterilize.

Phenol (pH 8.0):Chloroform(1:1): For 1L: In a chemical fume hood, adjust the pH of the phenol with the tris buffer included with it. Close the lid and shake vigorously. Mix 500 ml of the phenol pH 8.0 and 500 ml of chloroform in a 1L glass bottle with a lid. Shake the mixture vigorously and let it separate overnight at 4˚C. Store at 4˚C for up to a month.

Protocol

I. Yeast Cell Culture and crosslinking. *Crosslinking is particularly critical to standardize for all samples since the output of this method is the rate of crosslinking.*

1) Grow cells in conditions appropriate for the particular strain in use and for the particular experiment for 2-3 doubling times to obtain

mid-log phase growing cells. Usually, a 200ml culture produces enough cells to make two or three 3C libraries.

2) Add 37% formaldehyde to a final concentration of 3% in the culture media.
3) Shake culture for 20mins at 25˚C at 200 rpm.
4) Quench crosslinking by adding 2X the volume of formaldehyde used of 2.5M Glycine.
5) Shake culture for 5mins at 25˚C at 200 rpm.
6) Spin cells at 1,800x *g* in a 200ml conical centrifuge tube for 5mins in table top centrifuge.
7) Pour off media and wash cells in 100 ml sterile water by pipetting the cells up and down until the cell pellet is re-suspended.
8) Spin cells again at 1,800x *g* for 5mins in table top centrifuge.
9) Pour off supernatent and re-suspend the cells in 5ml of 1X restriction enzyme buffer by pipetting the cells up and down. *The restriction enzyme buffer used should be appropriate for the chosen restriction enzyme. Enzymes which rely on NEBuffer 4 should be avoided since the acetate ions precipitate the SDS in later steps.*
10) Cool a mortar and pestle by placing them on dry ice and add enough liquid nitrogen to the mortar to cover the head of the pestle. Let the liquid nitrogen evaporate.

11) Add another amount of liquid nitrogen to the mortar and pour the sample into the liquid nitrogen.
12) Once the sample has frozen, begin to crush it with the pestle. When the sample is broken into little pieces begin to grind it with the pestle. Grind the sample for 10mins, adding liquid nitrogen as necessary (approximately every 3mins).
13) Scrape the sample into a 50ml conical tube on ice.
14) Add 45ml of cold 1X restriction enzyme buffer to the lysate.
15) Spin the lysate at 1,800x *g* for 5mins in a table top centrifuge at 4˚C.
16) Pour off the supernatent.
17) Adjust the OD_{600} to 10.0 with 1X restriction enzyme buffer. *The lysate can be stored in 5ml aliquots at -80˚C for several years.*

II. Restriction Enzyme Digestion of Crosslinked Chromatin

1) Thaw one 5 ml aliquot of lysate on ice.
2) Wash the lysate with 45 ml of cold 1X restriction enzyme buffer in a 50 ml spin tube by inverting the tube several times.
3) Spin lysate at 3,100x *g* for 10 min in table top centrifuge at 4˚C.
4) Aspirate supernatent with a vacuum.
5) Re-suspend the pellet in 3.8 ml of 1X restriction enzyme buffer.
6) Distribute 38 µl of cells into 96 wells across one 96 well PCR plate.

7) Solubilize the chromatin by adding 3.8 µl 1% w/v SDS per well. Mix by pipetting up and down but avoid making bubbles.
8) Incubate in a thermocycler for 10 min at 65˚C. *Place the plate on ice immediately after the incubation since high temperature reverses formaldehyde crosslinks.*
9) Quench the SDS by adding 4.4 µl of 10% v/v Triton X-100 per well. Mix by pipetting up and down but avoid making bubbles.
10) Digest the chromatin by adding 5 µl of restriction enzyme (20,000 units/ml) per well. Mix by pipetting up and down but avoid making bubbles.
11) Incubate at 37˚C overnight in a thermocycler.
12) Denature the restriction enzyme by adding 8.6 µl of 10% SDS to every well. Mix by pipetting up and down but avoid making bubbles.
13) Incubate in a thermocycler for 20mins at 65˚C. *Place the plate on ice immediately after the incubation since high temperature reverses formaldehyde crosslinks.*

III. Intra-molecular Ligation of Crosslinked Chromatin Fragments. *It is essential to standardize this step for all samples since variations in ligation conditions can contribute to variation in background ligation.*

1) Set up the ligation reaction by adding the following to a 200ml flask on ice: 7.15 ml of 10% v/v Triton X-100, 7.15 ml of 10X Ligation

Buffer, 768 µl of 10 mg/ml BSA, 768 µl of 100 mM ATP, 57.2 ml of sterile water, 5.74 ml (combined) of the 96 3C digestion reactions, and 384 µl of T4 DNA Ligase. Mix the reaction by gently swirling but avoid making bubbles.

2) Split the ligation reaction into 8X 15 ml conical tubes (~9.6 ml per tube) on ice.
3) Incubate all ligation reactions at 16˚C for 4 hr.

IV. Reverse Crosslinking and Purification of Ligation Products

1) To every ligation reaction add 60 µl of 10mg/ml proteinase K. Mix by inverting.
2) Incubate all reactions at 65˚C for 4 hr.
3) To every ligation reaction add another 60 µl of 10mg/ml proteinase K. Mix by inverting.
4) Incubate all reactions at 65˚C overnight.
5) Add each reaction to a 50 ml conical tube.
6) Add 19.8 ml of phenol (pH 8.0):chloroform (1:1) to each tube.
7) Vortex each tube for 30 sec.
8) Pour the vortexed sample into a pre-spun 50 ml Phase Lock Light tube.
9) Centrifuge at 1,500x *g* for 10 min.
10) Pour the aqueous phase into a fresh 50 ml conical tube.
11) Add 19.8 ml of phenol (pH 8.0):chloroform (1:1) to each tube.

12) Vortex each tube for 30 sec.

13) Pour the vortexed sample into a pre-spun 50 ml Phase Lock Light tube.

14) Centrifuge at 1,500x *g* for 10 min.

15) Pool all 3C aqueous phases into 2X 250ml screw-cap centrifuge tubes suitable for high speed spinning (~39.6 ml each)

16) Precipitate the 3C ligation products by adding 1/10th volumes (3.96 ml) of 3M sodium acetate pH 5.2 and 2.5x volumes (99.0 ml) 100% ethanol to each tube.

17) Incubate all precipitation on dry Ice for 30-45 min. Make sure the liquid is very cold and thick but not frozen.

18) Centrifuge all precipitations at 10,000x *g* for 20 min at 4˚C.

19) Carefully pour off the supernatent.

20) Centrifuge all precipitations at 10,00 x *g* for 30 sec at 4˚C to collect the remaining drops of liquid at the bottom of the tube.

21) Aspirate the remaining alcohol using a vacuum.

22) Re-suspend both 3C DNA pellets in 8 ml total of 1X TE and add it to a 50 ml centrifuge tube.

23) Add 2 volumes (16 ml) of phenol (pH 8.0):chloroform (1:1) to the 3C sample.

24) Vortex each tube for 30sec.

25) Add the vortex 3C sample to a pre-spun 50 ml Phase Lock Light tube.
26) Centrifuge the 3C sample at 1,500x *g* for 10 min in a table top centrifuge.
27) Remove the aqueous phase from the 3C sample by pouring it into a clean 50 ml conical tube.
28) Add 2 volumes (16 ml) of phenol (pH 8.0):chloroform (1:1) to the 3C sample.
29) Vortex the tube for 30sec.
30) Add the vortex 3C sample to a pre-spun 50 ml Phase Lock Light tube.
31) Centrifuge the 3C sample at 1,500x *g* for 10mins in a table top centrifuge.
32) Remove the aqueous phase from the 3C sample and place it in a 35 ml centrifuge tube, suitable for high speed spinning.
33) Precipitate the 3C DNA with 1/10th volume (800 µl) sodium acetate pH 5.2 and 2.5x volume (20 ml) 100% ethanol.
34) Incubate on dry ice until the precipitation mixture is thick but not frozen
35) Centrifuge the 3C sample at 18,000x *g* for 20mins at 4˚C.
36) Decant the supernatent and let the DNA pellet air dry at room temperature.

37) Resuspend the 3C sample in 15 ml of 1X TE.

38) Desalt the sample by adding it to a 15 ml Amicon 30kda column.

39) Spin the 3C sample at 3,100x*g* in a table top centrifuge for 15 min.

40) Discard the flow through and wash the 3C sample 1 additional time with 15 ml of 1X TE.

41) Spin the 3C sample at 3,100x*g* in a table top centrifuge for 15 min

42) Elute the 3C from the 15 ml Amicon column by pipetting the remaining sample out of the filter with a 200 µl micro-pipette and put it in a clean 1.7 ml micro-centrifuge tube.

43) Degrade any co-precipitated RNA by adding 1/10th volume of 10mg/ml RNase A

44) Incubate at 37˚C for 1 hr.

45) Quantify the DNA using the method of choice and run on a 0.8% agarose gel to verify the 3C products (**Fig. 3.2c**). *The 3C products should run around 12kb on the gel as a tight band. Any lower molecular weight smearing would indicate a poor quality library and is probably due to over vigorous lysis.*

46) Prior to conducting an end-point PCR experiment with this 3C library one should titrate the PCR reaction to determine the amount of 3C library one needs to add to obtain quantitative PCR signals. Begin with 10ng of input 3C library and dilute it 8-10 by cutting the concentration in half. Run PCR reactions with primer will detect

interactions from a variety of genomic distances. For instance use a primer pair which will detect a long range interaction (100 – 150kb), a pair which will detect a medium range interaction (50 – 70kb), and a pair which will detect a close range interaction (10 – 20kb). Quantify the gel band products and plot the quantities graphically. Do this for all 3C libraries which will be used in a given experiment. Pick an amount of 3C library which is in the linear range of detection for all three distances in all of the 3C libraries and use this amount of 3C library for PCR with all primer-pairs in the 3C experiment.

Randomized Ligation Control for Chromosome Conformation Capture

Whether doing 3C or 5C it is critical to control for biases which are intrinsic to the restriction fragments of interest and the probes or primers used to detect those fragments (Dekker 2006; Naumova et al. 2012). These biases include GC%, primer annealing temperature (Tm) and efficiency of 3C primers or 5C probes, length of the restriction fragments etc. (Yaffe and Tanay 2011a; Imakaev et al. 2012). These characteristics of the fragments can cause a deviation in the performance of the 3C PCR primers, the 5C probes or the ligation efficiency of the fragment. To empirically measure this bias, a random control library is produced. This library is made by first purifying genomic DNA from cells and

then digesting the naked DNA with the same restriction enzyme used in the 3C experiment. Contrary to the 3C protocol, ligation of these restriction fragments is done in a concentrated condition that maximizes the number of random inter-molecular combinations of chimeric ligation products formed. The ligation product library generated after purifying these fragments is used for either 3C-PCR or 5C in the normal fashion. The signal produced from any combination of 3C primers or 5C probes is assumed to be the same since all biological signal has been removed from the data. Therefore, any deviation in the signal can be attributed to experimental biases of the fragments being interrogated and the probes/primers being used. The experimental 3C/5C data can then be corrected for such biases using these empirically measured deviations.

Materials

Reagents

1X Restriction Enzyme Buffer (which matches the chosen restriction enzyme)

10X Ligation Buffer

1X TE

Agarose

Appropriate growth media for the yeast strain of interest

ATP (10 mM)

Bovine Serum Albumin (BSA)(1 mg/ml)

EDTA (500 mM, pH 8.0)

Ethanol (100%)

Isopropanol (100%)

Lysing Buffer I

Phenol (pH 8.0):Chloroform (1:1)

Potassium acetate (5M)

Proteinase K (20 mg/ml in 1X TE)

Restriction Endonuclease of choice

RNase A (10mg/ml)

Sodium Acetate (3M, pH 5.2)

Spheroplasting Buffer II

Sterile Water

T4 DNA Ligase (1U/µl Life Technologies, 15224-017)

Recipes

10X Ligation Buffer: *Store at -20˚C in 15ml aliquots.*

Stock Solution	Amount for 1L	Final Concentration
1 M Tris-HCl pH 7.5	500 ml	500 mM Tris-HCl pH 7.5
1 M $MgCl_2$	100 ml	100 mM $MgCl_2$
1 M Dithiothreitol	100 ml	100 mM Dithiothreitol
De-ionized water	To 1L	

1X TE:

Stock Solution	Amount for 1L	Final Concentration
1 M Tris-HCl pH 8.0	10 ml	10 mM Tris-HCl pH 8.0
0.5 M EDTA pH 8.0	2 ml	1mM EDTA pH 8.0
De-ionized water	To 1L	

Lysing Buffer I: *make fresh for each experiment.*

Stock Solution	Amount for 100 ml	Final Concentration
0.5 M EDTA pH 8.0	50 ml	0.25 M EDTA
1.5 M tris-base	33.3 ml	0.5 M Tris-base
15% (v/v) SDS	16.7 ml	2.5% (v/v) SDS

Spheroplasting buffer II: *make fresh for each experiment.*

Stock Solution	Amount for 100 ml	Final Concentration
100 mM Sodium Phosphate buffer pH 7.2	10 ml	10 mM Sodium Phosphate buffer pH 7.2
500 mM EDTA	2.0 µl	10 mM EDTA
Beta-mercaptoethanol	1 ml	1% (v/v) beta-mercaptoethanol
20 mg/mL Zymolyase 100-T	0.5 ml	100 µg/mL zymolyase 100-T

Zymolyase 100-T solution (20 mg/ml): *Store at 4˚C for up to 1 month.*

Stock Solution	Amount for 1ml	Final Concentration
Zymolyase 100-T	20mg	20 mg/mL Zymolyase 100-T
40% Glucose	50.0 µl	2% (w/v) glucose
1 M Tris pH 7.5	50.0 µl	*50 mM Tris pH 7.5*
Sterile Water	To 1 ml	

3M Sodium Acetate pH5.2 (1L): Dissolve 246.1g of sodium acetate in 500ml of de-ionized water. Adjust the pH to 5.2 with Glacial Acetic Acid. Allow the solution to cool overnight. Adjust the pH once more to 5.2 with Glacial Acetic Acid. Adjust the final volume to 1L and filter sterilize.

Phenol (pH 8.0):Chloroform(1:1): For 1L: In a chemical fume hood, adjust the pH of the phenol with the tris buffer included with it. Close the lid and

shake vigorously. Mix 500 ml of the phenol pH 8.0 and 500 ml of chloroform in a 1L glass bottle with a lid. Shake the mixture vigorously and let it separate overnight at 4˚C. Store at 4˚C for up to a month.

Equipment

15 ml conical tubes

16˚C water bath or thermomixer

1.7 ml microcentrifuge tubes

2 ml Phase Lock Light tubes (5prime, 2302840)

200 ml Conical Centrifuge tubes

250 ml Screw cap tubes for high speed spinning.

35 ml Screw cap tubes for high speed spinning.

37˚C water bath or thermomixer

4˚C refrigerator

50 ml conical tubes

50 ml Phase Lock Light tubes (5prime, 2302860)

500 µl Amicon 30KDa column

65˚C water bath or thermomixer

High Speed Centrifuge (up to 18,000xg)

Spectrometer to measure the Optical Density at 600nm

Table top centrifuge capable of spinning 3,100x *g*

Thermocycler

Variable temperature incubator for growing yeast

Variable temperature shaking incubator for growing yeast

Vortex

Protocol

I. Isolate Yeast Chromosomal DNA

1) Obtain an overnight 200 mL culture of *Saccharomyces cerevisiae* using appropriate growth conditions.
2) Spin the cells for 10 minutes at 3,100x *g* in a 200 mL disposable conical centrifuge tube and remove supernatant promptly.
3) Resuspend cells in 20 mL spheroplasting buffer II and place it in a 35 ml centrifuge tube for high speed spinning.
4) Incubate for 40 minutes at 37˚C. Solution should appear stringy.
5) Add 4 ml lysing buffer I to the tube.

6) Add 400 µL 20 mg/mL proteinase K in TE buffer, pH 8.0 to the tube and incubate for 30 minutes at 65˚C. Solution should now appear clearer.
7) Add 4 ml of 5 M potassium acetate to each tube and incubate for 10 minutes in ice water.
8) Pellet the cell debris by centrifuging the tubes for 20 minutes at 18,000x *g* at 4˚C.
9) Split the supernatants in to 4 fresh 35 ml spin tubes suitable for high speed spinning (7.1 ml in each).
10) To each tube add 16.4 ml of ice cold 100% ethanol (70% final volume). Invert tubes 5 times to mix.
11) Precipitate the DNA by centrifuging the sample for 10 minutes at 18,000x *g* at 4˚C.
12) Remove supernatant carefully and let DNA pellets dry completely.
13) Resuspend all 4 DNA pellets in a total of 20ml of 1X TE buffer, pH 8.0 containing 10 µg/mL DNase-free RNase A in a 50 ml conical tube.
14) Incubate for 30 minutes at 37˚C to degrade the RNA. Occasionally tap tubes gently to mix.
15) Add 20.0 ml (1 volume) of phenol pH 8.0:chloroform (1:1)
16) Vortex the mixture for 30 seconds then add to a pre-spun 50 ml phase lock light tube.

17) Centrifuge at 1,500x *g* for 10 min to separate the organic and aqueous phases.
18) Transfer aqueous (upper) phase to 2 fresh 35 ml centrifuge tubes suitable for high speed spinning (10.0 ml in each).
19) Precipitate DNA by adding of 10ml of isopropanol to each tube and invert the tubes 5 times to mix. Centrifuge for 10 minutes at 18,000x *g* at room temperature. *DNA should be visible after isopropanol is added as a string- like ball.*
20) Remove the supernatant from the DNA pellet.
21) Wash the DNA pellet by resuspending it with 20 ml 80% ethanol with mild agitation and centrifuge 10 minutes at 18,00 x *g* at room temperature.
22) Remove the supernatant from the DNA pellet and allow it to dry at room temperature.
23) Resuspend the DNA pellet in 2ml 1X TE buffer.
24) Determine DNA concentration using absorption spectroscopy.

II. Genomic DNA Digestion

1) In 20x 1.7 ml microfuge tubes add 10.0 µg genomic DNA, 40.0 µl of 10X restriction enzyme buffer, 60 units of restriction enzyme, and sterile water to 400.0 µl to each.
2) Incubate all 20 reactions at 37˚C for 3 hours.

3) Pool all 20 reactions together in a 50 ml conical tube. The total volume will be 8.0 ml.
4) Extract the restriction enzyme from the reactions by adding 16.0 ml of phenol (pH 8.0)/chloroform (1:1).
5) Vortex the extraction mixture for 30 sec.
6) Add the vortexed extraction mixture to a pre-spun 50 ml Phase Lock Light tube.
7) Centrifuge the 3C sample at 1,500x *g* for 10 min in a table top centrifuge.
8) Pour the upper aqueous phase into a clean 35 ml tube suitable for high speed centrifugation.
9) Precipitate the DNA by adding 800.0 µl (1/10th volume) of 3M Sodium Acetate pH 5.2 and 20.0 ml (2.5 volumes) of 100% ethanol and mix by inversion.
10) Incubate the precipitation at -80˚C until the solution becomes thick but not frozen (~1hr).
11) Centrifuge the sample for 10 minutes at 18,000 x *g* at 4˚C.
12) Pour off the supernatant and let the DNA pellet dry at room temperature.
13) Re-suspend the DNA pellet in 15 ml of sterile water and vortex to dissolve the pellet.

14) Add the sample to a 15ml Amicon 30kDa molecular weight cut off column and spin in a table top centrifuge at 3,100x *g* until the sample is concentrated to 400 µl.

III. Random Ligation of Digested Genomic Restriction Fragments

1) In 20x, 1.7 ml microfuge tubes add 20.0 µl of Digested DNA, 3.0µl of 10X Ligation buffer, 3.0 µl 1 mg/ml BSA, 3.0 µl 10mM ATP and 1.0 µl of T4 DNA Ligase.
2) Incubate the reaction for 1hr at 16˚C.
3) Pool all 20 reactions (600µl) in a 2ml microfuge tube.
4) Stop the reaction by adding 12.0 µl of 0.5 M EDTA, pH 8.0.
5) Add 1.2 ml (2 volumes) of phenol pH 8.0:chloroform (1:1) and vortex for 30 seconds.
6) Split the extraction mixture into 2 pre-spun 2ml phase lock light tubes (900 µl each).
7) Spin the tubes at 18,000x *g* in a microcentrifuge.
8) Remove the upper aqueous phase with a micropipette and put each of them in a fresh 1.7 ml microcentrifuge tube (300µl).
9) Precipitate the DNA in each of the tubes by adding 30.0 µl of 3M Sodium acetate, pH 5.2 and 750 µl of 100% ethanol.
10) Mix by inversion, then spin in a micro-centrifuge for 10 min at 18,000x *g* at 4˚C.

11) Aspirate the supernatant with a vacuum or pipette and let the DNA pellet air dry at room temperature.

12) When the pellets are dry, resuspend both of them in a total of 500 µl of 1X TE.

13) Wash out any excess salt by placing the sample in a 500 µl Amicon 30kDa molecular weight cutoff column and spinning in a microcentrifuge for 5mins at 18,000x *g*. Bring the volume up to 500 µl with 1X TE and spin again at 18,000x *g*. Repeat the wash once more for a total of two washes.

14) Adjust the final volume to 400 µl with 1X TE.

15) Quantify the DNA with the method of choice.

16) Prior to conducting an end-point PCR experiment with this control library one should titrate the PCR reaction. Begin with 100ng of input control library and dilute it 8-10 by cutting the concentration in half. Run PCR reactions with primer will detect interactions from a variety of genomic distances. For instance use a primer pair which will detect a long range interaction (100 – 150kb), a pair which will detect a medium range interaction (50 – 70kb), and a pair which will detect a close range interaction (10 – 20kb). Quantify the gel band products and plot the quantities graphically. All of the titration curves should be very similar since there is no longer any biology represented in this control library. Pick an amount of control library

which is in the linear range of detection for all three distances and use this amount of control library for PCR with all primer-pairs in the 3C experiment. Quantify the gel band signal from both the 3C PCR and the control PCR for the same primer-pair and plot it as a ratio of the 3C signal divided by the control signal. This will normalize the each individual PCR for its own intrinsic biases discussed above.

Chromosome Conformation Capture Carbon Copy (5C) in Budding Yeast

Chromosome Conformation Capture Carbon Copy (5C) is a high throughput method to detect 3C ligation products of interest (Dostie et al. 2006a). 5C utilizes the technique Ligation Mediated Amplification (LMA) and literally makes a carbon copy of 3C ligation product junctions using short single stranded oligonucleotide probes (**Fig. 3.3b**). These 5C probes are designed adjacent to restriction fragments cut sites. 5C employs two types of probes: forward and reverse 5C probes. Forward probes anneal to the very end of the - strand of a restriction fragments of interest, whereas reverse probes anneal to the + strand of another restriction fragment of interest. For each fragment either a forward or a reverse 5C probe is designed. 5C can only detect interactions between forward and reverse probes: A pair of forward and reverse 5C probes can anneal to a head-to-head ligated pair of restriction fragments in a 3C ligation library.

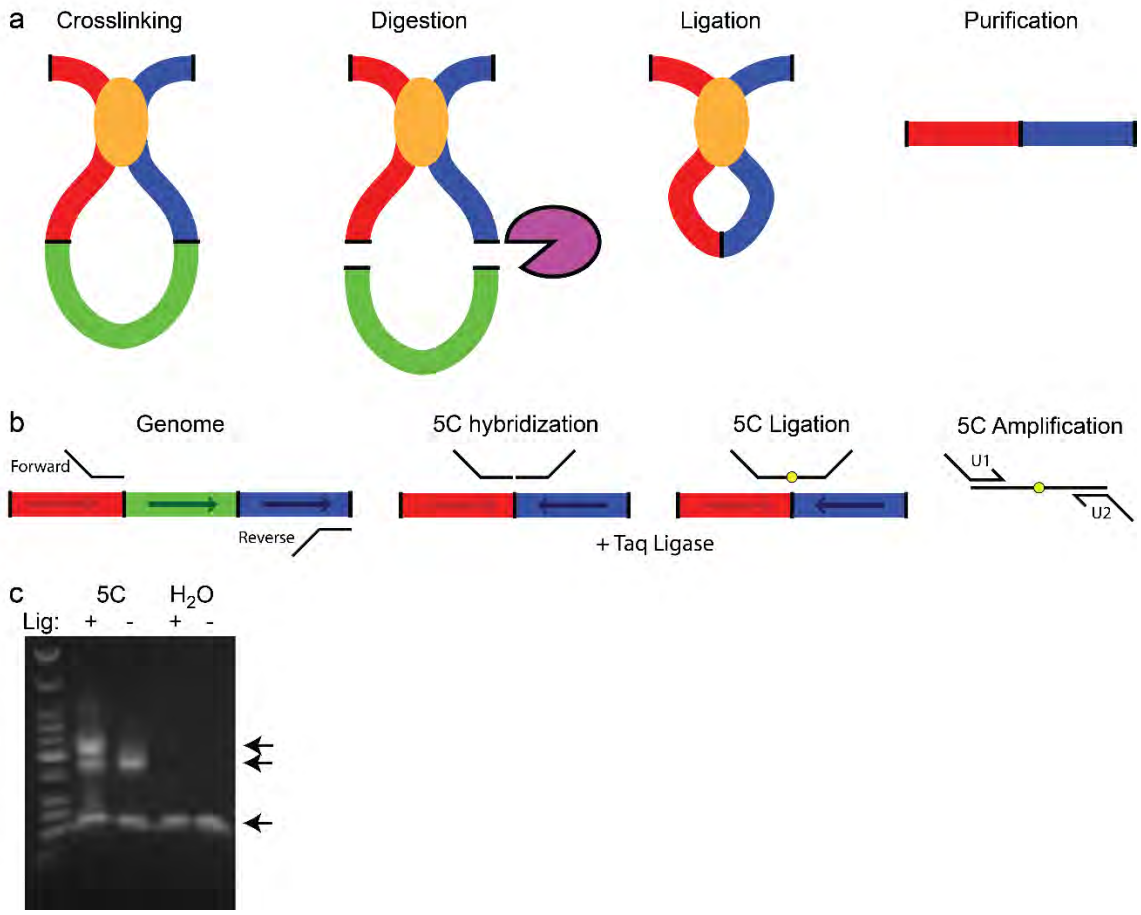


Figure 3.3. Outline of 5C. (a) The first step of 5C is to perform 3C to generate a 3C library that reflect relative interaction frequencies between pairs of restriction fragments. In 3C the chromatin is crosslinked with formaldehyde, digested with a restriction enzyme, then crosslinked fragments are ligated in an intra-molecular fashion. After purification of the DNA one can measure how frequently two fragments interacted with each other spatially by quantifying the resulting library of chimeric molecules. (b) 5C uses Ligation Mediated Amplification (LMA) to make copies of 3C ligation junctions. This process can be done at a high level of multiplexing, with hundreds to thousands of 5C probes, so that hundreds of thousands of chromatin interactions can be quantified in a single reaction. The 5C probes are designed to the genome directly adjacent to restriction fragment cut sites of the fragments of interest. There are two types of probes, forward and reverse. Forward probes hybridize to the "-" strand while reverse probes hybridize to the "+" strand. This arrangement prevents detection of ligation products formed by self-ligation of partial digestion products. The LMA reaction works by mixing the 5C probes with the 3C library. The 3C library is denatured and the 5C probes are allowed to anneal to 3C molecules. In cases where two

restriction fragments of interest are ligated together in the 3C library then the two respective 5C probes will anneal adjacent to each other. Taq Ligase is used to repair the nick between the 5C probes, creating a continuous molecule. Universal primers are then used to amplify the 5C ligation products and to incorporate sequences at the tails to facilitate deep sequencing. The molecules can then be sequenced with the platform of choice following the manufacturer's instructions. (c) 5C products and controls run on a gel. The 5C + Taq Ligase reaction has two bands. The correct 5C product is the top band (upper arrow). The middle arrow indicates the extended reverse probes which are present irrespective of the Taq Ligase. The bottom arrow indicates the left over universal primers from the PCR reaction. The exact size of the 5C library will depend on the size of the probes designed. The marker is the NEB low Molecular Weight Marker (New England Biolabs, N3233S).

This will leave a nick between the 5C probes that can be ligated using the nick-specific Taq ligase. As a result, a forward and reverse 5C probe are ligated and this ligation product is then PCR amplified using universal primers that contain the Illumina PE sequences. This copying process preserves the information contained in the junction between two restriction fragments but reduces the size of the molecules and modifies their tails to facilitate high throughput DNA sequencing. 5C can be carried out with a high degree of multiplexing which allows for the simultaneous detection of hundreds of thousands of ligation products at once in a single tube. This procedure produces a 5C library of short DNA molecules which represent the interaction between the corresponding restriction fragments and that can be sequenced with high throughput DNA sequencing methods. The advantages of using 5C to detect 3C ligation products of interest are that 1) 5C increases the speed of data production compared to doing 3C-PCR for the same number of interactions, 2) 5C can be used to assay many more interactions of interest than is practical with 3C-PCR, 3) The dynamic range of quantification of the 5C products is much longer than that of semi-quantitative 3C-PCR, 4) for very large and comprehensive studies the cost per interaction is much less than using 3C-PCR.

Materials

Reagents

10X NEBuffer 4 (New England Biolabs, B7004S)

10X PNK Buffer (New England Biolabs, M0201S)

10X Taq DNA Ligase Buffer (New England Biolabs, M0208S)

3C Library

5C probes

ATP (10 mM)

dNTP's (100 mM)

Pfu Ultra II HS enzyme (Agilent Technologies, 600850)

Sterile Water (de-ionized and autoclaved)

T4 Polynucleotide Kinase (New England Biolabs, M0201S)

Taq DNA Ligase (New England Biolabs, M0208S)

Universal PCR Primer 1: 5' -
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC
TCTTCCGATCTCCTCTCTATGGGCAGTCGGTGAT- 3' (HPLC purified)

Universal PCR Primer 2: 5' -
CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAA
CCGCTCTTCCGATCTCTGCCCCGGGTTCCTCATTCTCT- 3' (HPLC
purified)

Equipment

1.7 ml Microcentrifuge tubes

15 ml conical tube

37˚C Water bath or thermomixer

65˚C Water bath or thermomixer

8 well PCR strip tubes 100 µl

96 well PCR plates

Agarose Gel running apparatus

Face shield

Scalpel

Thermocycler

UV light box

Protocol

I. Phosphorylation of reverse probes and mixing of all 5C probes
    1) Mix all reverse probes in an equimolar ratio in a 1.7ml microfuge tube.
    2) In a separate 1.7ml microfuge tube mix all of the forward probes in an equimolar ratio.

3) Dilute the forward probe mixture so that the final concentration will be 0.2 fmol of each individual forward probe per µl.
4) To a 1.7 ml tube add 300 pmol of reverse probe mix, 1.0 µl (10 units) of T4 Polynucleotide Kinase, 5.0 µl 10mM ATP, 5.0 µl 10X PNK Buffer in a final volume of 50.0 µl.
5) Incubate the reaction at 37˚C for 45 mins.
6) Inactivate the enzyme by incubating the reaction at 65˚C for 20 mins.
7) Dilute the reverse probe mixture so that the final concentration will be 0.2fmol of each individual reverse probe per µl.
8) Mix equal proportions of the diluted forward probe mix and the phosphorylated reverse probe mix which will make the final concentration 0.1fmol of each probe per µl. *Large batches of this primer pool can be made at once then aliquotted and stored at -20˚C for several years.*

II. Annealing Reaction

1) In a PCR strip tube make 6x 5C reactions by adding 4.0µl of 10X NEBuffer 4, 10.0 µl of 5.33 ng/µl 3C library (4 million genome copies), 10.0 µl of 0.1fmol/µl probe mixture and 16.0 µl of sterile water for each reaction.
2) In the strip tube set up 2x water controls by adding 4.0µl of 10X NEB4 buffer, and 36.0 µl of sterile water for each reaction.

3) In a thermocycler perform the following program on all reactions.
    i. Denature the 3C library by incubating at 95˚C for 9mins.
    ii. Slowly decrease the temperature to 55˚C at a 0.1˚C/sec ramp. *If the probes are not made with the recommendations below then the annealing temperature will need to be determined for that specific set of probes.*
    iii. Incubate overnight at 55˚C.

III. Ligation

1) In a 1.7ml tube make the "+ligase" master mix by adding 3.0 µl of Taq DNA Ligase, 24.0 µl 10X Taq DNA Ligase Buffer, 213.0 µl of sterile water.
2) In another 1.7ml tube make the "-ligase" master mix by adding 8.0 µl 10X Taq DNA Ligase Buffer, 72.0 µl of sterile water.
3) Mix both of the Master mixes by vortexing.
4) While the 5C reactions are still at 55˚C add 40µl of the "+ligase" master mix to 5 of the 5C reactions and to one of the water controls. Add 40µl of the "-ligase" master mix to the last 5C reaction and to the last water control.
5) Incubate the reaction for 1 hour at 55˚C.

IV. 5C PCR amplification with Universal primers.

1) To make the PCR master mix, add 200.0 µl of 10X Pfu Ultra II buffer, 40.0 µl of 80 µM universal primer 1, 40.0 µl of 80 µM

universal primer 2, 16.0 µl of 100mM dNTP's, 1184.0 µl of sterile water, 40.0µl of Pfu Ultra II. *There are many choices for the universal sequence that are used to amplify the libraries. These sequences could be custom made or could be the adapter sequences for a specific sequencing platform. If sequences are chosen that are for modifying the library for a specific sequencing platform then this would save time and money preparing the library for sequencing. The advantage of having a sequence that is not for a particular platform is that different sequencing platforms can be used throughout the life of the probe set. For this protocol we use the universal primers which anneal to the 5C products but have the Illumina Paired-end sequences hanging off the end. Therefore, after amplifying the 5C library the Illumina sequences are already incorporated into the molecule.*

2) In a PCR plate take 12.0 µl of the 5C +Taq Ligase reaction and mix it with 38.0µl of PCR master mix. Use the entire volume of the 5C +Taq Ligase reactions (~30 reactions). Only make one PCR reaction for each of the three controls.
3) In a thermocycler run the following program: 1) 95˚C for 9 min, 2) 95˚C for 30 sec, 3) 65˚C for 30 sec, 4) 72˚C for 45 sec, 5) Go to step 2) 29 more times, 6) 72˚C for 7 min, 7) 4˚C forever.
4) Pool all of the 5C + Taq Ligase reactions together in a 15.0 ml tube.

5) Run three dilutions of the 5C + Taq Ligase reactions along with 10µl of each of the three controls on a 2% agarose gel.
6) Quantify the 5C bands on the gel and compare to a known standard run on the same gel to get an estimate of the concentration of the 5C library (**Fig. 3.3c**). *There will be two bands in the "5C + Ligase" lane. The lower band is the reverse probes that have bound to their universal primer and have polymerized the DNA to make these probes double stranded. This band will show up in the "5C - Ligase" control also. The other larger molecular weight band is the 5C library. This band will not show up in the "5C - Ligase" control indicating that it is dependent on the Taq ligase. The size of this band should be the size of 2 probes plus the Illumina PE adaptors. Only quantify the full length product. No product should be observed in either of the water controls.*
7) Take the remainder of the 5C + Taq Ligase reaction and run it on a long 2% agarose gel at 4˚C for 3hrs. *This will make sure that the bands will be spaced out enough so that it will be easy to cut out the full length 5C product.*
8) Using a scalpel and a face shield, cut out the larger full length band from the gel.
9) Purify this DNA product using any preferred Gel extraction procedure.

10) Quantify the 5C product using the method of choice and sequence the sample.

5C Probe Design

There are two types of 5C probes: forward and reverse. Both types of probes are adjacent to restriction cut sites. Forward probes hybridize to the "-" strand and have the universal sequence on the 5' end of the molecule. Reverse probes hybridize to the "+" strand and have the universal sequence on the 3' end of the molecule. This eliminates false positive signal arising from self-ligation of partial digestion products. There are many ways 5C probes can be positioned in the genome. Probes can be positioned such that forward and reverse probes alternate at consecutive restriction fragments. This layout will give the most density of structural information and therefore is ideal for structural/modeling studies of chromatin (Bau and Marti-Renom 2011; Bau et al. 2011). Also, probes can be placed in a manner in which the two types of probes represent different types of chromatin classes. For instance one type of probe could represent gene promoters and another would represent enhances (Sanyal et al. 2012).

5C probes should be designed to be 40 nt long (excluding universal tail sequences) and have an optimal Tm of 65˚C. The probes should also hybridize to unique sequences in the genome to avoid any off target effects. The Web based tool: http://my5c.umassmed.edu can be used to design 5C probes (Lajoie et al. 2009a).

Yield of 5C Reaction

The yield of a 5C reaction will depend on how the 5C probes are arranged in the genome. For instance a probe arrangement which detects only inter-chromosomal interactions will yield fewer products since there are fewer of these products in a 3C library compared to a similar design which would detect only intra-chromosomal interactions. The ratio of 4 million genome copies to 1.0 fmol of each primer per 5C reaction is optimized for detecting a mixture of both inter-chromosomal and intra-chromosomal interactions. If the yield of the five 5C reactions mentioned in the protocol does not produce enough DNA for downstream processing then one can simply make more 5C reactions with the same conditions and then pool all reactions together.

Hi-C in Budding Yeast

Hi-C is a technique that allows for the simultaneous detection of interaction frequencies between all possible pairs of restriction fragments in the genome (Lieberman-Aiden et al. 2009a; Belton et al. 2012). Hi-C is based on Chromosome Conformation Capture (3C)(Dekker et al. 2002b) and uses formaldehyde crosslinking to fix chromatin regions that are physically interacting in three-dimensional space, irrespective of their genomic locations (**Fig. 3.4a**). Cross-linked chromatin is digested with a restriction enzyme and then the restriction cut site is filled in with a nucleotide mix containing biotinylated dCTP.

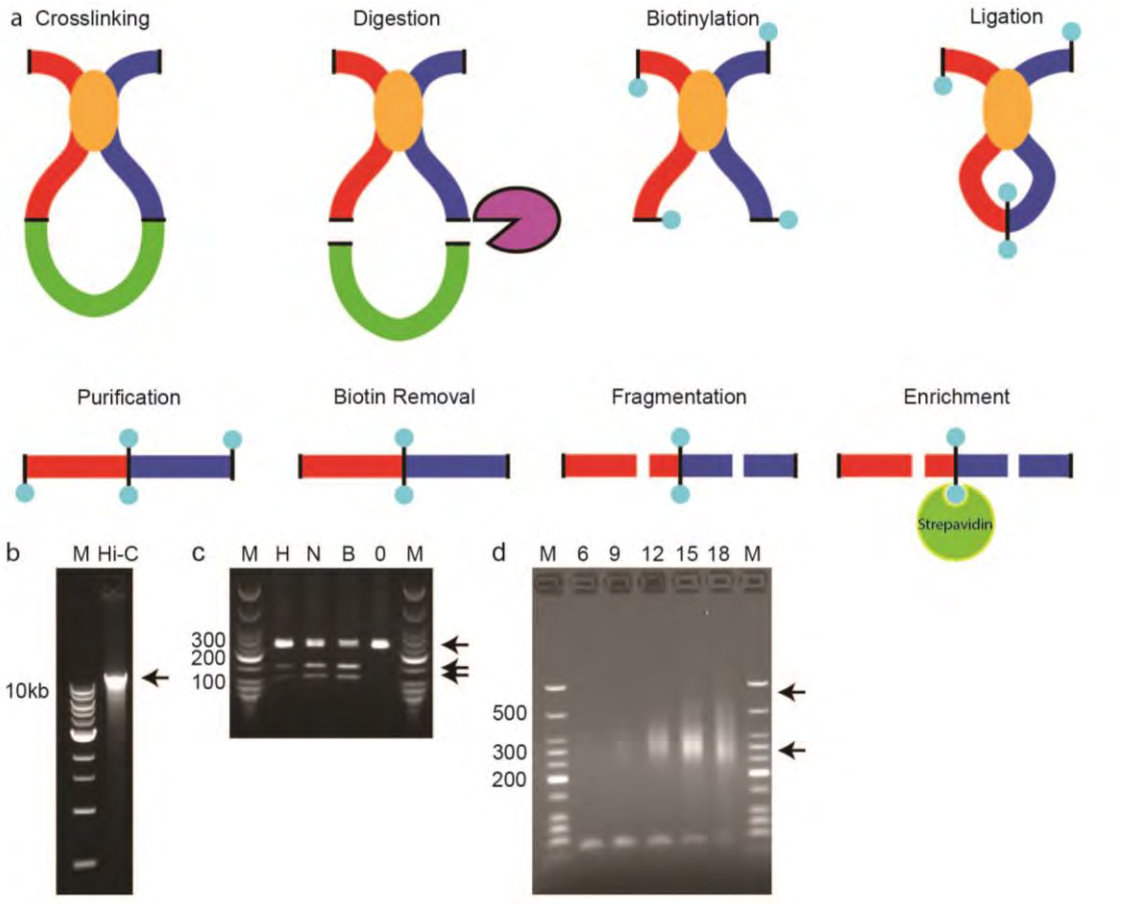


Figure 3.4. Schematic overview of the Hi-C procedure and quality control steps. Hi-C is very similar to 3C in that the chromatin is crosslinked using formaldehyde and digested with a restriction enzyme. The difference is that the overhangs of the restriction cut site are filled in with a biotinylated nucleotide prior to ligation. (a) Intra-molecular ligation and purification are also very similar to the 3C protocol. Hi-C has a few additional steps to enrich for ligation products. First, biotin that was incorporated into free ends (not ligation junctions from two different fragments) is removed using the exonuclease activity of T4 DNA polymerase. The library is then fragmented by sonication and then true ligation products are enriched by pulling-down the biotinylated ligation products using streptavidin coated beads. (b) The purified Hi-C library from section V step 51 should run at ~12kb (arrow). The marker is a 1 kb DNA ladder (New England Biolabs, N3232S). (c) The Hi-C efficiency gel from section VI step 4. “H” is HindIII digested, “N” is NheI digested, “B” is Both HindIII and NheI digested, and “0” is no digestion of the PCR product. “M” is the low molecular weight marker (New England Biolabs, N3233S). The upper arrow indicates the undigested PCR product and the two lower arrows indicate the digestion products. (d) PCR cycle

titration from section XII step 5. “M” is the low molecular weight marker (New England Biolabs, N3233S). “6”, “9”, “12”, “15”, “18” are the number of cycles used to amplify the library. The bottom arrow is the correct size of the Hi-C library and the upper arrow indicates high molecular weight species that arise from over amplifying the library.

The restriction fragments that are cross-linked to one another are ligated together in a dilute reaction in order to favor intra-molecular ligation. These chimeric molecules are then purified and sheared to reduce the length of the molecules. Biotinylated ligation junctions are pulled down with streptavidin-coated beads and the molecules are linkered with high throughput sequencing adaptors. This method provides the most comprehensive genome-wide analysis of any of the 3C-based techniques. The resolution of the dataset will depend on the depth of sequencing, and the choice of restriction enzyme. When sufficient sequence reads are obtained, information on chromatin interactions and chromosome conformation can be derived at single restriction fragment resolution for complete genomes.

Materials

Reagents

1X Binding Buffer (1xBB)

1x NEBuffer 2 (New England Biolabs, B7002S)

1X TE

1X TLE

1X Tween Wash Buffer (TWB)

10X Ligation Buffer

10X Pfu Ultra II HS buffer (Agilent Technologies, 600850)

2X Binding Buffer (2xBB)

Ampure XP (Beckman Coulter, Beckman Coulter)

Appropriate growth media for the yeast strain of interest

ATP (100 mM)

biotin-14-dCTP (0.4 mM, Life Technologies, 19518-018)

Bovine Serum Albumin (BSA)(10 mg/ ml)

dATP (100 mM)

dGTP (100 mM)

dNTP mix (25mM each nucleotide, 100mM total)

DNA Polymerase I, Large (Klenow) Fragment (5U/µl, New England Biolabs, M0210S)

Dry ice

dTTP (100 mM)

Dynabeads MyOne Strepavidin C1 beads (Life Technologies, 65001)

Ethanol (100%)

EB (Qiagen, 19086)

Formaldehyde (37%)

Glycine (2.5M, filter sterilized)

HindIII (New England Biolabs, R0104S)

Illumina PE adapters (Illumina, PN 1001784)

Klenow Fragment (3’ → 5’ exo-)(5U/µl, New England Biolabs, M0212S)

NEB Quick Ligation Kit (New England Biolabs, M2200S)

NheI (New England Biolabs, R0131S)

Oligonucleotide: Hi5, GTTTCCGAAAATCCACGACGAACCAG

Oligonucleotide: Hi6, ATATTTTCGCCGGAGGTGCTGGAAAT

Oligonucleotide: PE1.0,
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC
TCTTCCGATCT (HPLC purified)

Oligonucleotide: PE2.0,
CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAA
CCGCTCTTCCGATCT (HPLC purified)

Pfu Ultra II HS enzyme (Agilent Technologies, 600850)

Phase Lock Light tubes (15 ml, 5Prime, 2302840)

Phase Lock Light tubes (50 ml, 5Prime, 2302860)

Phenol (pH 8.0):Chloroform (1:1)

Proteinase K (10 mg/ ml in 1X TE, Life Technologies, 25530-015)

RNase A (10mg/ ml)

Sodium Acetate (3M, pH 5.2)

Sodium Dodecyl Sulfate (1%)

Sodium Dodecyl Sulfate (10%)

Sterile Water (de-ionized, and autoclaved)

T4 DNA Ligase (1U/µl, Life Technologies, 15224)

T4 DNA Polymerase (3U/µl, New England Biolabs, M0203L)

T4 Polynucleotide Kinase (10U/µl, New England Biolabs, M0201)

Triton X-100 (10%)

Recipes

1X TE:

Stock Solution	Amount for 1L	Final Concentration

1 M Tris-HCl pH 8.0	10 ml	10 mM Tris-HCl pH 8.0
0.5 M EDTA pH 8.0	2 ml	1mM EDTA pH 8.0
De-ionized water	To 1L	

1X TLE:

Stock Solution	Amount for 1L	Final Concentration
1 M Tris-HCl pH 8.0	10 ml	10 mM Tris-HCl pH 8.0
0.5 M EDTA pH 8.0	200 µl	0.1mM EDTA pH 8.0
De-ionized water	To 1L	

1X Binding Buffer (1X BB):

Stock Solution	Amount for 100 ml	Final Concentration
1 M Tris-HCl pH8.0	500 µl	5.0 mM Tris-HCl pH8.0
500 mM EDTA	100 µl	0.5 mM EDTA
5 M NaCl	20 ml	1 M NaCl
De-ionized water	To 100 ml	

1X Tween wash Buffer (1X TWB):

Stock Solution	Amount for 100 ml	Final Concentration
1 M Tris-HCl pH8.0	500 µl	5.0 mM Tris-HCl pH8.0
500 mM EDTA	100 µl	0.5 mM EDTA
5 M NaCl	20 ml	1 M NaCl
100% Tween 20	50.0 µl	0.05% Tween 20
De-ionized water	To 100 ml	

10X Ligation Buffer: *Store at -20˚C in 15 ml aliquots.*

Stock Solution	Amount for 1L	Final Concentration
1 M Tris-HCl pH 7.5	500 ml	500 mM Tris-HCl pH 7.5
1 M $MgCl_2$	100 ml	100 mM $MgCl_2$
1 M Dithiothreitol	100 ml	100 mM Dithiothreitol
De-ionized water	To 1L	

2X Binding Buffer (2X BB):

Stock Solution	Amount for 100 ml	Final Concentration
1 M Tris-HCl pH8.0	1.0 ml	10 mM Tris-HCl pH8.0
500 mM EDTA	200 µl	1 mM EDTA

5 M NaCl	40 ml	2 M NaCl
De-ionized water	To 100 ml	

3M Sodium Acetate pH5.2 (1L): Dissolve 246.1g of Sodium Acetate in 500 ml of de-ionized water. Adjust the pH to 5.2 with Glacial Acetic Acid. Allow the solution to cool overnight. Adjust the pH once more to 5.2 with Glacial Acetic Acid. Adjust the final volume to 1L with de-ionized water and filter sterilize.

Phenol (pH 8.0):Chloroform(1:1): For 1L: In a chemical fume hood, adjust the pH of the phenol with the tris buffer included with it. Close the lid and shake vigorously. Mix 500 ml of the phenol pH 8.0 and 500 ml of chloroform in a 1L glass bottle with a lid. Shake the mixture vigorously and let it separate overnight at 4˚C. Store at 4˚C for up to a month.

Equipment

15 ml Amicon 30KDa column (Millipore, UFC903024)

15 ml conical tubes

16˚C water bath or thermomixer

1.7 ml LoBind tubes

1.7 ml microcentrifuge tubes

250 ml Screw cap tubes for high speed spinning.

35 ml Screw cap tubes for high speed spinning.

37˚C water bath or thermomixer

50 ml conical tubes

500 µl Amicon 30KDa column (Millipore, UFC5030BK)

65˚C water bath or thermomixer

Covaris micro tube AFA fiber with snap-cap (Covaris, 520045)

Covaris S2 (Covaris)

Large centrifuge capable of spinning 35 ml and 250 ml tubes at 18,000x *g*

Magnetic Particle Separator (MPS)

Mortar and pestle

Qiagen MinElute Columns (Qiagen, 28004)

Rocking platform

Table top centrifuge capable of spinning 50 ml conical tubes at 3,100x *g*

Thermocycler

Variable temperature incubator for growing yeast

Variable temperature shaking incubator for growing yeast

Vortex

Protocol

I. Yeast Cell Culture and crosslinking. *Crosslinking is particularly critical to standardize for all samples since the output of this method is the rate of crosslinking itself.*

1) Grow cells in conditions appropriate for the particular strain in use and for the particular experiment for 2-3 doubling times to obtain mid-log phase growing cells. Usually, a 100 ml culture produces enough cells to make one Hi-C library.
2) Add 37% formaldehyde to a final concentration of 3% in the culture media (8.82 ml of 37% formaldehyde for 100 ml culture).
3) Shake culture for 20 min at 25˚C at 200 rpm.
4) Quench crosslinking by adding 2X the volume of formaldehyde used of 2.5M Glycine (17.64 ml of 2.5M Glycine for 100 ml culture).
5) Shake culture for 5 min at 25˚C at 200 rpm.
6) Spin cells at 1,800x *g* in a 250 ml conical centrifuge tube for 5 min in table top centrifuge at room temperature.
7) Pour off media and wash cells in 100 ml sterile water by pipetting the cells up and down until the cell pellet is re-suspended.

8) Spin cells again at 1,800x *g* for 5 min in table top centrifuge.
9) Pour off supernatant and re-suspend the cells in 5 ml of 1X NEBuffer 2 by pipetting the cells up and down.
10) Cool a mortar and pestle by placing them on dry ice and adding enough liquid nitrogen to cover the head of the pestle. Let the liquid nitrogen evaporate.
11) Add more liquid nitrogen to the mortar and pour the sample into the liquid nitrogen.
12) Once the sample has frozen, begin to crush it with the pestle. When the sample is broken into little pieces begin to grind it with the pestle. Grind the sample for 10 min, adding a little liquid nitrogen as need (approximately every 3 min) to keep the sample cool.
13) Scrape the sample into a 50 ml conical tube on ice.
14) Add 45 ml of cold 1X NEBuffer 2 to the lysate.
15) Spin lysate at 1,800x *g* for 5 min in a table top centrifuge at 4˚C.
16) Pour off the supernatant.
17) Adjust the OD_{600} to 10.0 with 1X NEBuffer 2. *The lysate can be stored in 7.2 ml aliquots at -80˚C for several years.*

II. HindIII Digestion of Crosslinked Chromatin

1) Wash one, thawed, 7.2 ml aliquot of lysate with 50 ml of cold 1X NEBuffer 2 in a 50 ml spin tube by inverting the tube several times.

2) Spin lysate at 3,100x *g* for 10 min in table top centrifuge at 4˚C.
3) Aspirate supernatant with a vacuum.
4) Re-suspend the pellet in 5.5 ml of 1X NEBuffer 2.
5) Distribute 456 µl of lysate into 12 x 1.7 ml microcentrifuge tubes.
6) Solubilize the chromatin by adding 45.6 µl of 1% w/v SDS per tube. Mix by pipetting up and down and re-suspend any cellular debris that has precipitated but avoid making bubbles.
7) Incubate in a 65˚C water bath or thermomixer for 10 min. *Place the tubes on ice immediately after the incubation since high temperature reverses formaldehyde crosslinks.*
8) Spin the tubes quickly for 10 sec. to remove liquid from the cap.
9) Quench the SDS by adding 52.8 µl of 10% v/v Triton X-100 per tube. Mix by pipetting up and down and re-suspend any cellular debris that has precipitated but avoid making bubbles.
10) Digest the chromatin by adding 60 µl of HindIII (20,000 units/ ml) per tube. Mix by pipetting up and down and re-suspend any cellular debris that has precipitated but avoid making bubbles.
11) Incubate at 37˚C overnight in a water bath or thermomixer. *It is preferable to agitate the sample during the incubation so a rotating platform in a 37˚C incubator is recommended.*

III. Filling-In and Biotin Incorporation of Digested Ends.

1) Spin the tubes quickly for 10 sec. to remove liquid from the cap.

2) To the Hi-C reactions add: 6.4 µl 10X NEBuffer 2, 0.18 µl 100mM dATP, 0.18 µl 100mM dGTP, 0.18 µl 100mM dTTP, 45.0 µl of 0.4mM biotin-14-dCTP, and 12.0 µl of 5U/µl Klenow.
3) Incubate at 37˚C for 2 hr in a water bath or thermomixer. *It is preferable to agitate the sample during the incubation so a rotating platform in a 37˚C incubator is recommended.*
4) Spin the tubes quickly for 10 sec. to remove liquid from the cap.
5) Denature the enzymes by adding 115.2 µl of 10% SDS to each tube. Mix by pipetting up and down but avoid making bubbles.
6) Incubate in a 65˚C water bath or thermomixer for 20 min. *Place the tubes on ice immediately after the incubation since high temperature reverses formaldehyde crosslinks.*

IV. Intra-molecular Ligation of Crosslinked Chromatin Fragments. *It is essential to standardize this step for all samples since variations in ligation conditions can contribute to variation in background ligation (inter-molecular ligation), which would reduce the signal to noise ratio.*

1) Add the following to a 250 ml flask on ice: 13.0 ml 10% v/v Triton X-100, 13.0 ml of 10X Ligation Buffer, 1.4 ml 10mg/ ml BSA, 1.4 ml 100mM ATP, 104.0 ml sterile water, and 3.51 ml T4 DNA Ligase. Mix the reaction by gently swirling but avoid making bubbles.
2) Add 10.5 ml of ligation master mix to each of 12 15 ml tubes on ice.

3) Spin the digestion tubes quickly for 10 sec. to remove liquid from the cap.
4) Using a 1 ml micropipette, remove each digestion reaction and add it to a ligation tube. Mix the tubes gently by inverting
5) Incubate the ligation reactions in a 16˚C water bath or thermomixer for 8 hr. Invert the tubes every hour to keep them mixed.

V. Reverse crosslinks and purification the ligation products

1) Add 72 µl of 10mg/ ml proteinase K to each ligation mixture. Mix by inverting gently.
2) Incubate the reactions in a 65˚C water bath or thermomixer overnight.
3) Add 72 µl of 10mg/ ml proteinase K to each ligation mixture. Mix by inverting gently.
4) Incubate the reactions in a 65˚C water bath or thermomixer for 2 hours.
5) Add each reaction to a 50 ml centrifuge tube.
6) Add 23.0 ml of phenol (pH 8.0):chloroform (1:1) per tube.
7) Vortex each tube for 30 sec.
8) Pour the vortexed sample into a pre-spun 50 ml Phase Lock Light tube.
9) Centrifuge at 1,500x *g* for 10 min.
10) Pour the aqueous phase into a fresh 50 ml centrifuge tube.

11) Add 23.0 ml of phenol (pH 8.0):chloroform (1:1) per tube.
12) Vortex each tube for 30 sec.
13) Pour the vortexed sample into a pre-spun 50 ml Phase Lock Light tube.
14) Centrifuge at 1,500x *g* for 10 min.
15) Pool all Hi-C aqueous phases into 2X 250 ml screw-cap centrifuge tubes suitable for high speed spinning (67.8 ml each)
16) Precipitate the DNA by adding 1/10th volumes (6.78 ml) of 3M sodium acetate pH 5.2 and 2.5x volumes (169.5 ml) of 100% ethanol to each tube.
17) Incubate all precipitations on dry ice for 30-45 min. Make sure the liquid is very cold and thick but not frozen.
18) Centrifuge all precipitations at 10,000 x *g* for 20 min at 4˚C.
19) Carefully pour off the supernatant.
20) Centrifuge all precipitations at 10,000x *g* for 30 sec at 4˚C to collect the remaining drops of liquid at the bottom of the tube.
21) Aspirate the remaining alcohol using a vacuum.
22) Re-suspend both Hi-C DNA pellets in 4 ml total of 1X TE and add it to a 15 ml centrifuge tube.
23) Add 8.0 ml of phenol (pH 8.0):chloroform (1:1) to the Hi-C.
24) Vortex each tube for 30sec.

25) Add the vortexed Hi-C sample to a pre-spun 15 ml Phase Lock Light tube.
26) Centrifuge the Hi-C sample at 1,500x *g* for 10 min in a table top centrifuge.
27) Remove the aqueous phase from the Hi-C sample and place in a 15 ml centrifuge tube.
28) Add 8.0 ml of phenol (pH 8.0):chloroform (1:1) to the Hi-C.
29) Vortex each tube for 30sec.
30) Add the vortexed Hi-C sample to a pre-spun 15 ml Phase Lock Light tube.
31) Centrifuge the Hi-C sample at 1,500x *g* for 10 min in a table top centrifuge.
32) Remove the aqueous phase from the Hi-C sample and place in a 35 ml centrifuge tube, suitable for high speed spinning.
33) Precipitate DNA with 400µl of sodium acetate pH 5.2 and 10.0 ml of 100% ethanol.
34) Centrifuge Hi-C sample at 18,000x *g* for 20 min.
35) Decant the supernatant.
36) Centrifuge all precipitations at 10,000 x *g* for 30 sec at 4˚C to collect the remaining drops of liquid at the bottom of the tube.
37) Aspirate the remaining alcohol using a vacuum.
38) Air dry the DNA pellet to remove any remaining ethanol.

39) Re-suspend the Hi-C sample in 15 ml of 1X TE and vortex for 60 sec.

40) Desalt the sample by adding the Hi-C sample to a 15 ml Amicon 30kda column.

41) Spin the Hi-C sample at 3,100x*g* in a table top centrifuge for 15 min.

42) Discard the flow through and wash the Hi-C sample 1 additional times with 15 ml of 1X TE.

43) Elute the Hi-C from the 15 ml Amicon column by pipetting the remaining sample out of the filter with a 200µl micro-pipette and put it on a 500µl Amicon 30kda column.

44) Wash the 15 ml Amicon column with an additional 200 µl of 1X TE and add to the rest of the sample in the 500µl Amicon column.

45) Spin the Hi-C sample at 18,000x *g* for 20 min. *If the volume in the column is greater than 20 µl then spin it repeatedly for 5 min at a time until it is less than 20 µl.*

46) Elute the Hi-C samples by inverting the column and placing it in a fresh collection tube.

47) Spin the Hi-C sample at 18,000x *g* for 20 min.

48) Adjusted the volume to 20 µl with 1X TE.

49) Degrade any co-precipitated RNA by adding 2 µl of 10mg/ ml DNase free RNase A

50) Incubate at 37˚C for 1 hr.

51) Quantify the library on a 0.8% agarose gel in 0.5X TBE by comparing it to known concentration standards (**Fig. 3.4b**). *The Hi-C library should run as a tight band around 12 kb. Any smearing of the sample may be due to overly vigorous lysis or the presence of endogenous nucleases in the strain. Typically 3 µg – 10 µg of Hi-C library are isolated from the sample with an average yield of approximately 5 µg.*

VI. Estimate Hi-C Efficiency. *The Hi-C efficiency is the percentage of ligation products for a single interaction that are biotinylated. This measurement is used to standardize the amount of streptavidin beads to use when enriching the library for biotinylated ligation products and gives a measure of how efficient the biotin fill-in was.*

1) PCR amplify a neighboring interaction in the Hi-C library.
    i. Make the PCR reaction by adding 40 ng of the Hi-C library, 10 µl of Pfu Ultra II buffer, 0.8 µl of 100mM dNTP (25mM each nucleotide), 0.5 µl of 80 µM primer Hi5, 0.5 µl of 80 µM primer Hi6, 2.0 µl of Pfu Ultra II, and sterile water to 100 µl total volume
    ii. Amplify the ligation product of interest using the following thermocycler program: 95˚C for 5 min, 95˚C for 30 sec, 65˚C for 30 sec, 72˚C for 30 sec, go to step 2 34x, 72˚C for 8 min.

2) Digest the amplicon with HindIII, NheI, and both HindIII and NheI.
    i. HindIII: 15µl PCR reaction, 1.9µl NEBuffer 2, 0.19µl BSA, 0.95µl HindIII (20,000 units/ ml), 0.95µl sterile water
    ii. NheI: 15µl PCR reaction, 1.9µl NEBuffer 2, 0.19µl BSA, 0.95µl NheI, 0.95µl sterile water
    iii. Both: 15µl PCR reaction, 1.9µl NEBuffer 2, 0.19µl BSA, 0.95µl HindIII (20,000 units/ ml), 0.95µl NheI
    iv. None: 15µl PCR reaction, 1.9µl NEBuffer 2, 0.19µl BSA, 1.9µl sterile water
3) Incubate the digestion reactions overnight at 37˚C.
4) Run the digestion reactions on a 2% agarose gel in 0.5x TBE and quantify the digested and undigested bands (**Fig. 3.4c**).
5) The Hi-C efficiency is the percentage of digested products in the NheI digestion. Not all PCR products will be cut in the NheI and HindIII combined reaction. This is likely due to inefficiency of the digestion and point mutations introduced during amplification and may also be due to the breakage of the DNA during the procedure resulting in non-canonical ligation products. The Hi-C efficiency should be corrected by the percent cleavable which can be determined from this reaction. Therefore, the final Hi-C efficiency is calculated as the percent digested with NheI divided by the percent digested in the NheI and HindIII combined reaction. *If the Hi-C*

efficiency is low this could be due to either a poor biotinylation reaction or a poor ligation reaction or both. Generally, 30-60% of the amplicon will digest with NheI.

VII. Removal of biotin from un-ligated, free ends. *Proceed with the entire Hi-C sample through the rest of the protocol.*

1) For each 1 µg of Hi-C library add: 0.1 µl of 10 mg/ml BSA, 1.0 µl of 10X NEBuffer 2, 0.1 µl of 5mM dNTP (1.25mM for each nucleotide), 1.0 µl of T4 DNA polymerase, and adjust the final volume to 10.0 µl with sterile water.
2) Incubate the reaction at 20˚C for 4hrs and then at 75˚C for 20 min to inactivate the enzyme.
3) Adjust the volume to 100 µl by either adding sterile water or concentrating the sample using the 500 µl 30kDa Amicon columns used previously.

VIII. Shearing of Hi-C library. *The Covaris S2 is the sonicator of choice since it shears the DNA to 50-700 bp fragments in only 4 min. Any other sonicator would also work but the exact conditions and setting would have to be determined empirically for naked DNA and may not produce as tight of a smear.*

1) Load 100 µl of the Hi-C sample into a Covaris microTUBE.

2) Fill the water chamber of the Covaris with deionized water and allow it to cool down and degas for at least 30 min prior to sonication.
3) Place the microTUBE into the Covaris approved S2 holder and run the following program: Duty Cycle 10%, Intensity 5, Cycles per Burst 200, Mode is Frequency sweeping. Continue to degas during the sonication. Run the program for a total of 4 min.
4) Purify the DNA with one Qiagen MinElute column for every 5µg of Hi-C reaction. Elute the DNA from the column with 31 µl of 65˚C EB.

IX. Repair of DNA ends and A-tailing

1) To the fractionated sample add: 14.0 µl 10X NEB Ligation Buffer, 1.4 µl of 100 mM dNTPs (25 mM each), 5.0 µl T4 DNA Polymerase, 5.0 µl T4 Polynucleotide Kinase, 1.0µl Klenow fragment of DNA Polymerase I, 13.6 µl sterile water.
2) Incubate the reaction at 20˚C for 30 min.
3) Purify the DNA with one Qiagen MinElute Column. Elute the DNA with 31 µl of 65˚C EB.
4) To the end-repaired sample add 5.0 µl of 10X NEBuffer 2, 10.0 µl of 1.0mM dATP, 3.0 µl of Klenow Fragment (exo-), 2.0 µl of sterile water.

5) Incubate the reaction at 37˚C for 30 min. Then inactivate the Klenow Fragment (exo-) by incubating the reaction at 65˚C for 20 min.

X. Ampure fractionation of the sonicated Hi-C library

1) Bring the volume up to 500µl with EB (Qiagen).
2) Use the Ampure XP mixture to bind DNA species with a molecular weight greater than 300bp by adding 450µl of Ampure XP mixture to the Hi-C sample. Label this sample “0.9X”
3) Vortex the mixture briefly then quickly spin the mixture to remove drops from the cap.
4) Incubate the mixture at room temperature for 10 min.
5) While the “0.9X” sample is incubating add 500 µl of Ampure XP to a new 1.7 ml tube and label this tube “1.1X”.
    i. Place the “1.1X” tube on a Magnetic Particle Separator (MPS) for 5min to collect the Ampure XP beads to the side of the tube.
    ii. While the tube is still on the MPS remove the Ampure solution using a 1 ml micropipette and discard.
    iii. Resuspend the collected beads in 100 µl of Ampure XP mixture. *This process increases the number of beads in the Ampure XP mixture and ensures that we will have enough binding capacity to collect all of the DNA in the sample.*

6) Place the “0.9X” tube on an MPS for 5 min to collect the Ampure XP beads to the side of the tube.
7) While the “0.9X” tube is still on the MPS remove the Ampure solution using a 1 ml micropipette and **add it to the “1.1X” tube**.
8) Vortex the “1.1X” tube briefly then quickly spin the mixture to remove drops from the cap.
9) Incubate the mixture at room temperature for 10 min.
10) Place the “1.1X” tube on an MPS for 5 min to collect the Ampure XP beads to the side of the tube.
11) While the “1.1X” tube is still on the MPS remove the Ampure solution using a 1 ml micropipette and discard.
12) Wash the “0.9X” and the “1.1X” beads twice with 200 µl of 70% ethanol. Do this by re-suspending the beads in the 70% ethanol, spinning the tube quickly and collecting the beads on the MPS.
13) After the final ethanol wash remove the remaining ethanol by spinning the sample once more and collecting left over ethanol. Then dry the beads at room temperature to allow the ethanol to evaporate.
14) Re-suspend both the “0.9X” and the “1.1X” beads in 30 µl of EB and incubate at room temperature for 10 min to elute the DNA from the Ampure beads.

15) Place both the “0.9X” and the “1.1X” tube on an MPS for 5 min to collect the Ampure XP beads to the side of the tube.
16) While both the “0.9X” and the “1.1X” tube are still on the MPS remove the Ampure solution using a 1 ml micropipette place in a new 1.7 ml tube. *The “0.9X” sample contains all DNA fragments that are greater than 300bp and the “1.1X” sample contains DNA fragments that are between 100bp and 300bp.*
17) Run the “un-sonicated”, “sonicated”, “0.9X” and “1.1X” samples on a short 2% agarose gel in 0.5X TBE at 250 volts for 30 min to verify the size of the DNA fragments and to quantify the 1.1X sample by comparing it to a known amount of a standard to verify the sizes.

XI. Enrichment of Biotinylated Ligation Products

1) Calculate the amount of Dynabeads MyOne Strepavidin C1 beads to use for the pull down:
    i. Multiply the amount of DNA measured in section IX step 17 by 200 (the average length of the sonicated Hi-C library) then divide this amount by 8000 (the average length of the pre-sonication library). This quantity is an estimate of the amount of the sample that contains a ligation product.
    ii. Next multiply the quantity by the proportion of biotinylated ligation products (Hi-C efficiency), which was calculated in

section VI step 5. This final quantity is the amount of DNA in the library which is a biotinylated ligation product.

2) For each 1.0 ng of biotinylated ligation products place 1.0 µl of Dynabeads MyOne Strepavidin C1 beads in to a 1.7 ml low binding tube. *Low binding tubes are used in this step to cut down the amount of non-specific pull-down of DNA.*
3) Wash the beads twice with 400 µl Tween Wash Buffer (TWB). *Each wash in this section consist of resuspending the beads in the appropriate solution using a micropipette, transferring the beads to a new low binding tube, incubating the beads for 3 min with rotation, quickly spinning the beads to the bottom of the tube, and reclaiming the beads using an MPS for 1 min.*
4) Re-suspend beads in 30 µl 2x Binding Buffer (BB).
5) Add the 30 µl of Hi-C library to the beads and incubate for 1 hour at room temperature with rotation.
6) Reclaim beads using the MPS for 1 minute. *Keep supernatant as insurance until the end of the protocol.*
7) Wash the beads once with 200 µl 1X BB.
8) Wash the beads once with 50 µl 1X NEB Quick Ligation Buffer.
9) Re-suspend the beads in 10 µl of sterile water.

10) Ligate the Illumina PE adapters by adding 13 µl of 2X NEB Quick Ligation Buffer, 2 µl of Illumina PE Adapters, and 1µl of NEB Quick Ligase.
11) Incubate at room temperature for 15 min.
12) Reclaim the beads using the MPS for 1 min.
13) Wash the beads twice with 200 µl 1X TWB.
14) Wash the beads once with 100 µl 1X BB.
15) Wash the beads once with 100 µl 1X NEBuffer 2.
16) Wash the beads once with 25 µl 1X NEBuffer 2.
17) Resuspend the beads in the amount of 1x NEBuffer 2 that is the same as the starting volume of beads used. *If 10 µl of Dynabeads MyOne Strepavidin C1 beads were used to pull down the library then resuspend the library in 10 µl of NEBuffer 2.*

XII. Paired-end PCR of Adapter Modified Hi-C library.

1) Titrate the number of PCR cycles to use to amplify the library by adding 3 µl of Hi-C coated streptavidin beads, 5 µl of Pfu Ultra II buffer, 0.4 µl of 100mM dNTPs (25mM each), 0.7 µl of PE PCR 1.0 primer, 0.7 µl of PE PCR 2.0 primer, 1µl Pfu Ultra II, and 39.2 µl sterile water.
2) Amplify the library using the following thermocycler program: 98˚C for 30 sec, 98˚C for 10 sec, 65˚C for 30 sec, 72˚C for 30 sec, go to step 2 5 times, 72˚C for 2 min.

3) Vortex the PCR reaction and remove 2.0 µl to run on a gel.
4) Quickly spin the reaction and repeat steps 2 and 3 except only run the PCR program for 3 cycles. Do this step until 18 total PCR cycles have been run on the sample.
5) Run all of the reactions on a short 2% agarose gel in 0.5X TBE and quantify the amount of DNA for each cycle number (**Fig. 3.4d**).
6) Pick a number of cycles that will yield at least 100 fmol of final Hi-C library when all of the beads are used but doesn't have higher molecular weight artifacts which arise due to over cycling. *Amplify all libraries in the same experiment with the same number of cycles. Use the least number of cycles possible and amplify all of the beads.*
7) Use the optimum number of cycles to PCR amplify the remaining beads using the reaction set up in step 1 and the thermocycler program in step 2.
8) After the PCR amplification, pool all reactions together and add 1.8 volumes of Ampure XP beads.
9) Place the tube on an MPS for 1 min to collect the beads.
10) Remove the supernatant and discard.
11) Wash the Ampure XP beads twice with 1 ml 70% ethanol while the tube remains on the magnet.
12) Air-dry the beads on the magnet.

13) Elute the DNA by re-suspending the beads in 14 µl of TLE.

14) Place the tube on an MPS for 1 min to collect the beads.

15) Remove the supernatant and place in a clean 1.7 ml tube.

16) Quantify the library using the method of choice (bioanalyzer, qPCR, fluorometry, etc…).

CHAPTER IV

The Recombination Enhancer Modulates the Conformation of Chr. III in Budding Yeast

Contributions

Jon-Matthew Belton performed all 5C and Hi-C analysis as well statistical analysis, Bryan R. Lajoie assisted with analysis, Job Dekker and Jon-Matthew Belton conceived the project and directed analysis. Kerstin Bystricky, Imen Lassadi and Isabelle Goiffon perform the live cell imaging.

Abstract

Yeast chromosome III contains the mating type loci that provide a paradigm for long-range interactions between distant loci. Yeast switches mating type by gene conversion between the *MAT* locus and either of two silent loci (*HML* or *HMR*) on opposite ends of the chromosome. This long-range process is mating type-specific so that *MAT***a** cells choose *HML* as a template, while *MAT*α cells use *HMR*. The Recombination Enhancer (RE) located on the left arm regulates this process. One long-standing hypothesis is that mating type switching is guided by mating type-specific, and possibly RE-dependent, three-dimensional folding of chromosome III. Here we used Hi-C, 5C, and live cell imaging to characterize the conformation of chromosome III in the two mating types. We discovered a mating-type specific difference in the folding of the left arm: in *MAT***a** cells the left arm is spatially close to the central centromere-containing portion of chromosome III, whereas it is more extended in *MAT*α cells. Surprisingly, deletion of the left part of the RE modestly affects the conformation of the left arm in *MAT*α cells, where the RE is in an inactive state. However the right part of the RE, which has a very modest effect on left arm usage in *MAT***a** cells during switching, is responsible for all of the conformational differences between the mating types. Interestingly, conformational changes are observed in both *MAT***a** and *MAT*α cells.

Introduction

Budding yeast has two mating types, *MAT***a** and *MAT*α**.** The mating type of a cell is determined by the *MAT* locus, located on the right arm of chromosome III. The *MAT* locus expresses the mating type specific transcription factors *MAT*α1p and *MAT*α2p in *MAT*α cells and *MAT***a**1p and *MAT***a**2p in *MAT***a** cells. These proteins are transcription factors responsible for establishment of the *MAT***a** or *MAT*α mating type. Besides the active *MAT* locus, chromosome III contains a silent copy of the *MAT*α locus located near the left telomere at *HML*α, and a silent copy of *MAT***a** near the right telomere at *HMR***a**. The silent copies of the *MAT* locus are important for the cell to switch its mating type. Mating type switching is initiated by the HO endonuclease that generates a DNA double strand break at the active *MAT* locus. This break is then repaired by a gene conversion reaction with either *HML*α or *HMR***a** as donor sequence (Haber 2012b).

The mating type switching process is highly directional. *MAT***a** cells strongly favor switching to the *MAT*α mating type by using *HML*α, whereas *MAT*α cells switch to the *MAT***a** mating type using *HMR***a**. Thus, the cell is able to direct the gene conversion reaction. The mechanisms for mating-type specific directionality of the gene conversion reaction at *MAT* are still only partially understood. Two phenomena are known to play roles in mating type switching. First, the left arm of chromosome III is unusually refractory to recombination, both for mating type switching and for gene conversion events in general (Wu and

Haber 1995b; Haber 2012b). Second, the Recombination Enhancer (RE), a cis-acting element located on the left arm, is required to activate the left arm for switching in *MAT***a** cells (Wu and Haber 1996a; Szeto et al. 1997b). The RE is composed of two highly conserved Mcm1p/*MAT*α2p binding sites called DPS1 and DPS2 and three highly conserved arrays of Fkh1 binding sites. The left 700bp portion of the RE contains the DPS1 operator and the three Fkh1 binding site arrays. The FHA domain of Fkh1 binds phosphor-serine and phospho-threonine residues that accumulate at double stranded break sites in *S. cerevisiae* (Li et al. 2012). It is thought that FKH1p binds to the RE and forms a physical bridge to the *MAT* locus after it has been cleaved by HO during mating type switching (Li et al. 2012). Switching assays have shown that this region as well as the binding of FKH1p is essential for the usage of the left arm during mating type switching in *MAT***a** cells, whereas the right-most ~1400bp, which includes DPS2, has only a modest effect on left arm usage in *MAT***a** cells (Wu and Haber 1996b; Szeto et al. 1997a).

In *MAT***a** cells, the RE is in an open chromatin state and binds Forkhead transcription factors that contribute to preferential activation of the left arm for the mating type switching reaction (Sun et al. 2002a). In *MAT*α cells, the RE is inactivated by binding of the Mcm1p/*MAT*α2p repressor complex that positions nucleosomes along the RE (Szeto et al. 1997b; Wu et al. 1998c).

The mechanisms that contribute to the general inaccessibility of the left arm for recombination and the mode of action of the RE in *MAT***a** cells are not

understood. One long-standing hypothesis is that the spatial conformation or sub-nuclear positioning of the left arm is preventing it from engaging in recombination. In this model the RE would re-position the left arm in *MAT***a** cells specifically to make it available for gene conversion with the *MAT* locus (Haber 1998a; Haber 2012b).

Here we have set out to determine the three-dimensional (3D) organization of chromosome III at 4-8 kb resolution by comprehensive mapping of long-range chromosomal interactions using Hi-C, 5C, and live cell imaging. We find that chromosome III displays a mating type-dependent spatial conformation, with the left arm more crumpled making it interact more frequently with the centromeric portion of the chromosome in *MAT***a** cells, and the left arm more extended from the centromere in *MAT*α cells. Interestingly, deletion of the left portion of the RE or of FKH1 affects the conformation only in *MAT*α cells, while deletion of the right portion of the RE affects the conformation in both *MAT***a** and *MAT*α cells. Our results firmly establish cell type-specific folding of the chromosome and provide new insights into the role of the RE on chromosome structure.

Results

Genome-wide Chromatin Interaction Maps for *MAT*a and *MAT*α Cells

To investigate the conformation of the yeast genome, we performed Hi-C on exponentially growing cultures of both *MAT***a** and *MAT*α cells as described

previously (Lieberman-Aiden et al. 2009a; Belton et al. 2012). Figure 4.1a shows genome-wide chromatin interactions maps for both cell types. Visual inspection of the heatmaps reveals that overall features of chromosome and nuclear organization are very similar in both mating types, and consistent with previous studies. First, both mating types display the characteristic Rabl configuration of chromosomes. In this conformation, all centromeres are clustered together at one side of the nucleus with the arms running in parallel towards the other end. In yeast this is known to be facilitated by tethering of centromeres to the spindle pole body, and the resulting Rabl arrangement has been directly observed by imaging as well as by 3C and previous genome-wide chromatin interaction analysis (Berger et al. 2008; Duan et al. 2010a) (Jin et al. 1998b; Jin et al. 2000b; Dekker et al. 2002b). In Hi-C interaction maps, centromere clustering and the Rabl arrangement are readily detectable by the presence of strong interactions between all pairs of centromeres (**Fig. 4.1a**). In addition, as a result of tethering to the spindle pole body, the chromosomes are kinked at the centromere with the pericentromeric parts of the arms running closely and in a parallel fashion alongside each other away from the nuclear periphery. This is

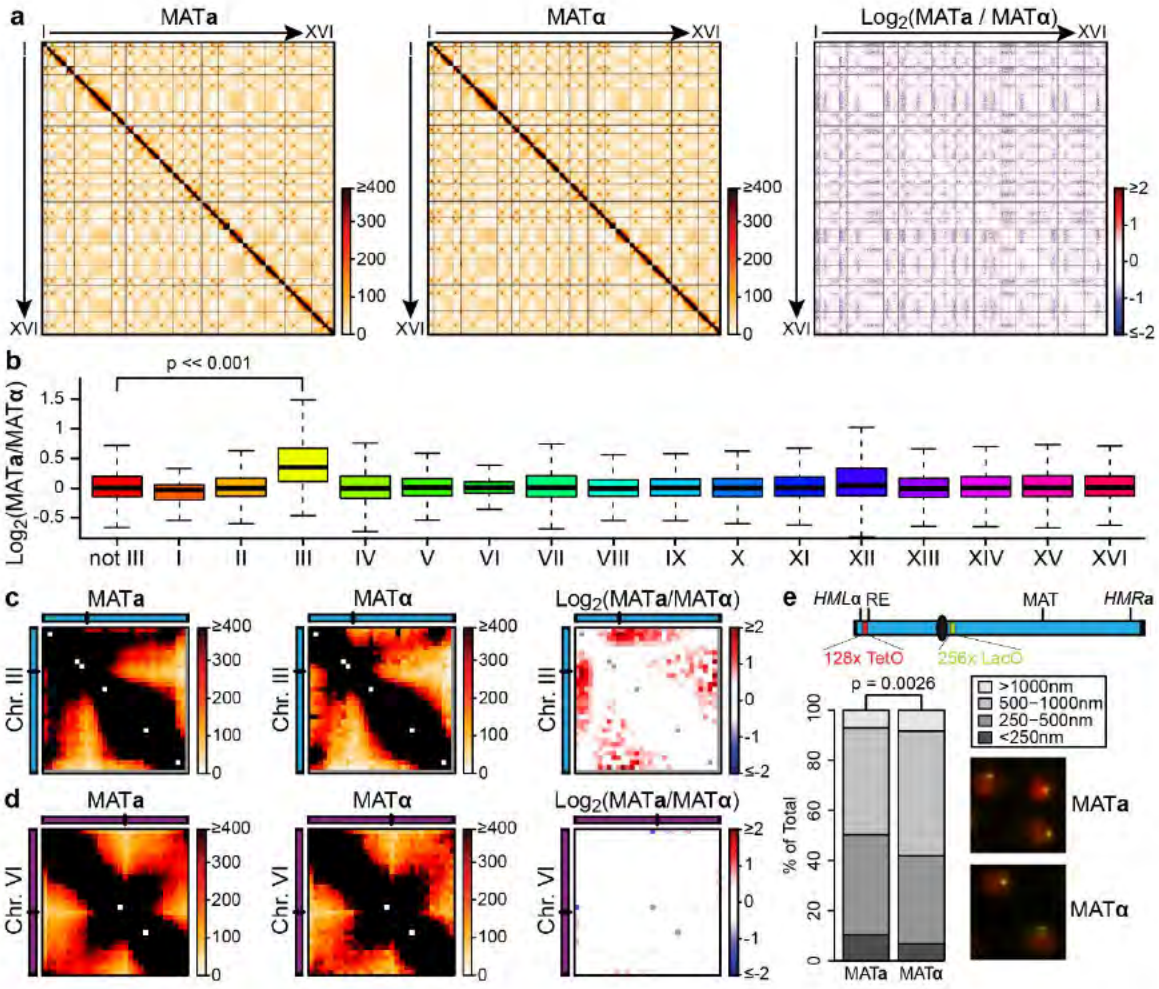


Figure 4.1. Global chromosome conformation in *MAT***a** and *MAT*α cells. (**a**) The left and middle panel are Hi-C from both *MAT***a** and *MAT*α cells respectively. Data from two biological replicates was pooled and corrected using the iterative coverage correction method at 10 kb resolution and the total number of counts has been normalized for each dataset. The chromosomes are sorted on both axes by the chromosome number. The right heatmap is the log2 ratio between the *MAT***a** and *MAT*α. (**b**) box plots of all the intra-chromosomal differences form the log2 comparison between *MAT***a** and *MAT*α. "not III" is all of the differences from all of the chromosomes except Chr. III. A rank sum test was used to calculate the p-value between Chr. III and "not III" (**c**) Zoom in to Chr. III. The data is the same as in (a). (**d**) Zoom in to Chr. VI. (**e**) the top shows a schematic of the bacterial operator insertions on chromosome III used to verify the Hi-C data. Below shows representative images for the two mating types and the bar graph indicates the percentage of cells that were in various distance categories.

visible in the Hi-C maps by the cross-shaped interaction patterns around each centromere (**Fig. 4.1c, d**). Similar cross-shaped patterns are present at each centromere-centromere interaction, indicating that the pericentric domains of all 32 arms run parallel to each other for up to 50 kb (Zimmer and Fabre 2011). A further indication of the Rabl organization is the depletion of interactions between centromeres and more distal portions of chromosome arms. This is driven by volume exclusion effects through which arms of each chromosome are forced to be extended out into the nucleus, reducing their probability of interacting with the centromere (Tjong et al. 2012). Additionally we also observe that short chromosome arms are enriched for interactions with other short arms but are depleted for interactions with long chromosome arms (**Fig. 4.2**). This has also been observed before (Duan et al. 2010a; Therizols et al. 2010) and is again due to a volume exclusion effect near the centromeres that pushes the long chromosomes further out into the nuclear volume (Tjong et al. 2012). Second, the telomeres interact with one another more frequently than would be expected for the genomic distance between them (Duan et al. 2010b; Agmon et al. 2013). This is consistent with observations that the telomeres are tethered to the nuclear periphery (Gotta et al. 1996; Trelles-Sticken et al. 2000; Berger et al. 2008; Bystricky et al. 2009a).

The folding of chromosome III is mating type dependent

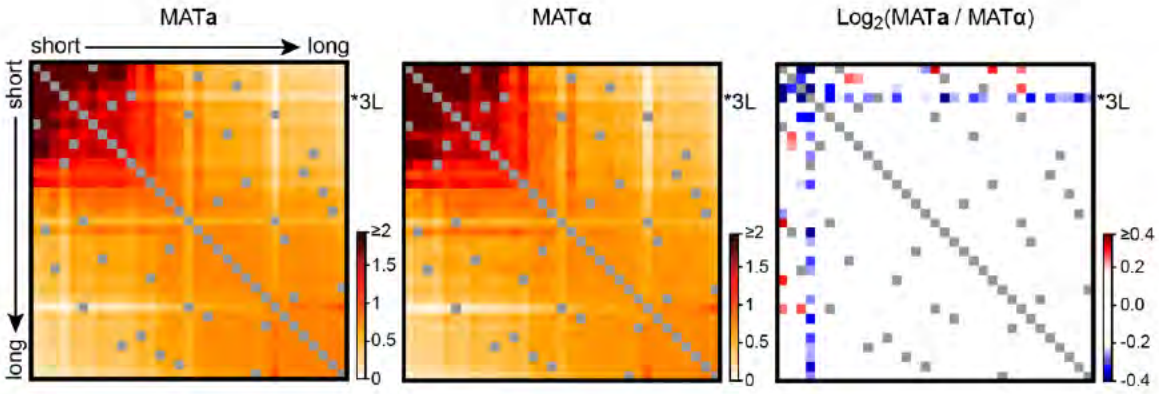


Figure 4.2. The left arm of Chr. III interacts more with other chromosome arms in *MAT*α cells. The left and middle heatmaps show the median interaction strength between all pairs of chromosome arms. The right panel is the log2 ratio of the left and middle heatmaps. Both the X and Y axes are sorted by arm length in all panels.

To compare the spatial organization of the genome in *MAT***a** and *MAT*α cells, we calculated the log2 ratio of the two Hi-C interaction maps and plotted the results again as a heatmap (**Fig. 4.1a right panel**). A weak global pattern of differences in inter-chromosomal centromere-arm interactions is observed that can be explained by a small differences in the noise level of very low level inter-chromosomal interactions between experiments and is observed in comparisons between biological replicates (**Fig. 4.3**) and is likely due to different levels of random ligation. This effect is not observed in intra-chromosomal interactions that are less susceptible to natural fluctuations of noise since the values are orders of magnitude higher than inter-chromosomal ones.

Comparison of interaction maps for each chromosome reveals that there are very few intra-chromosomal differences in conformation, with log ratios of differences centered around zero (**Fig. 4.1b**). Interestingly, chromosome III stands out as having the most differences between the mating types and is statistically significant compared to all of the differences from all of the other chromosomes combined. Therefore, we analyzed the conformation of chromosome III in more detail (**Fig. 4.1c**). In both of the mating types we observe general conformational features that are shared with other chromosomes: a general inverse relationship between interaction frequency and genomic distance and a cross-shaped pattern of interactions around the kinked centromere. In addition, in both mating types

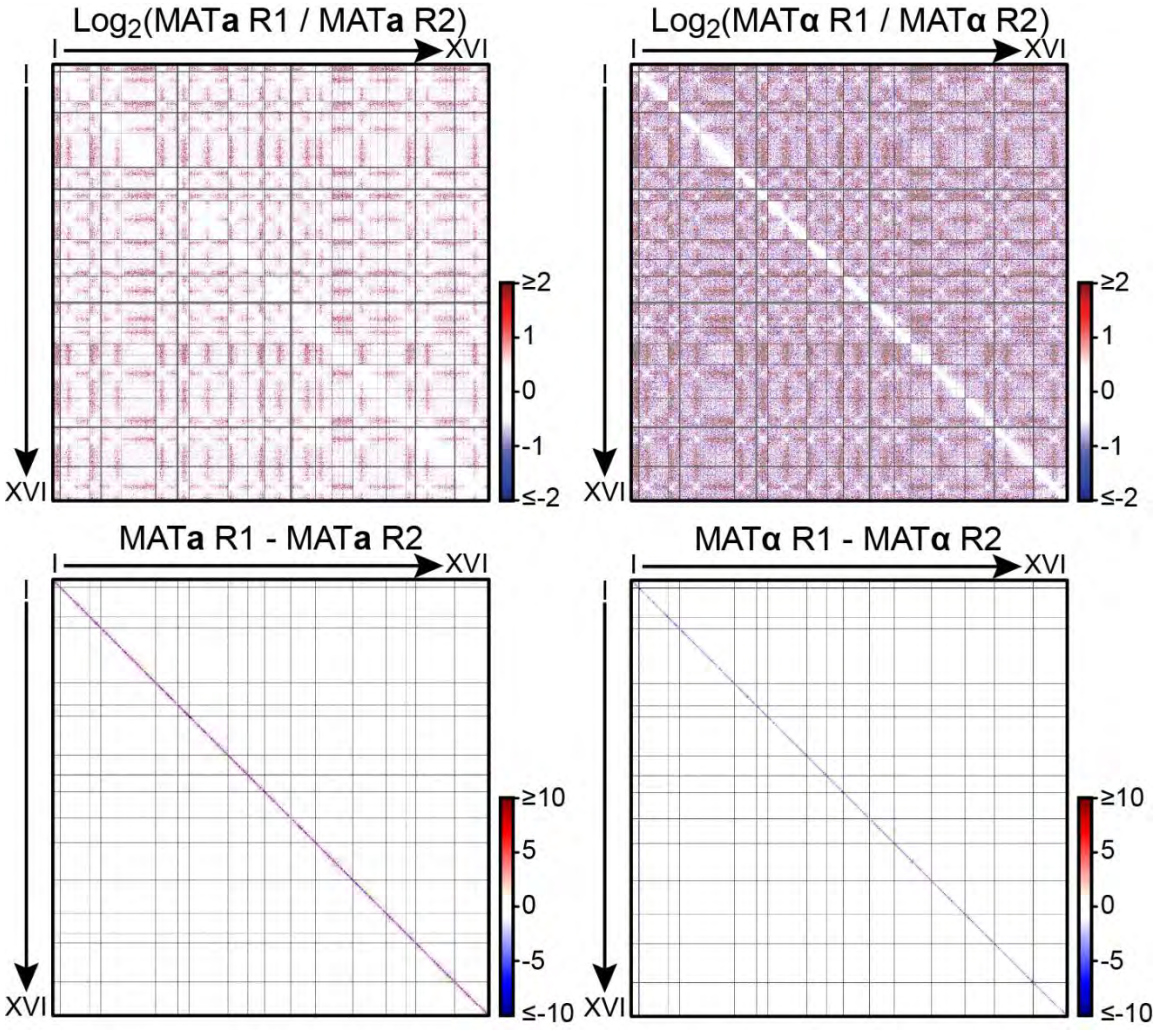


Figure 4.3. Differences in noise in Hi-C samples in inter-chromosomal regions. Top left is the log2 ratio between *MAT***a** biological replicates. Top right log2 ratio between *MAT*α replicates. Bottom left is the difference between *MAT***a** replicates and bottom right is the difference between *MAT*α replicates.

we observe prominent interactions between the heterochromatic hidden *MAT* loci (*HML* and *HMR*) located on opposite ends of the chromosome (**Fig. 4.1c**), which have been detected and characterized in detail previously (Dekker et al. 2002b; Bystricky et al. 2009b; Miele et al. 2009). These interactions are not merely due to the interactions between the telomeres of chromosome III because they are much stronger than interactions observed between telomeres on other small chromosomes such as chromosome VI (**Fig. 4.1d**) and because *HMR* is positioned far enough from the end of the right arm as to be distinguished from the telomere of that arm. Further, deletion of *HMR* has been shown to eliminate this long-range interaction (Miele 2009).

We wondered why *HML* and *HMR* interactions are important for the biology of Chr. III. We leveraged the massive amount of chromosomal interaction data from the Hi-C experiments to measure how often chromatin regions interact together given the spatial positioning of *HML*, *HMR*, and *MAT* without any additional mechanism. Figure 4.4 is a schematic of how the data for this analysis was collected. To generate the distribution of expected interactions between *HML* and *MAT* we collected interactions between two loci from short or medium length chromosomes positioned the same distance from the centromere as *HML* and *MAT*. To generate the expected distribution of interactions between *MAT* and *HMR*, we collected interactions that are on the same medium sized chromosome arm and are the same distance apart as *MAT* and *HMR* are from one another. We found that, given the genomic positioning of *HML*, *MAT*, and

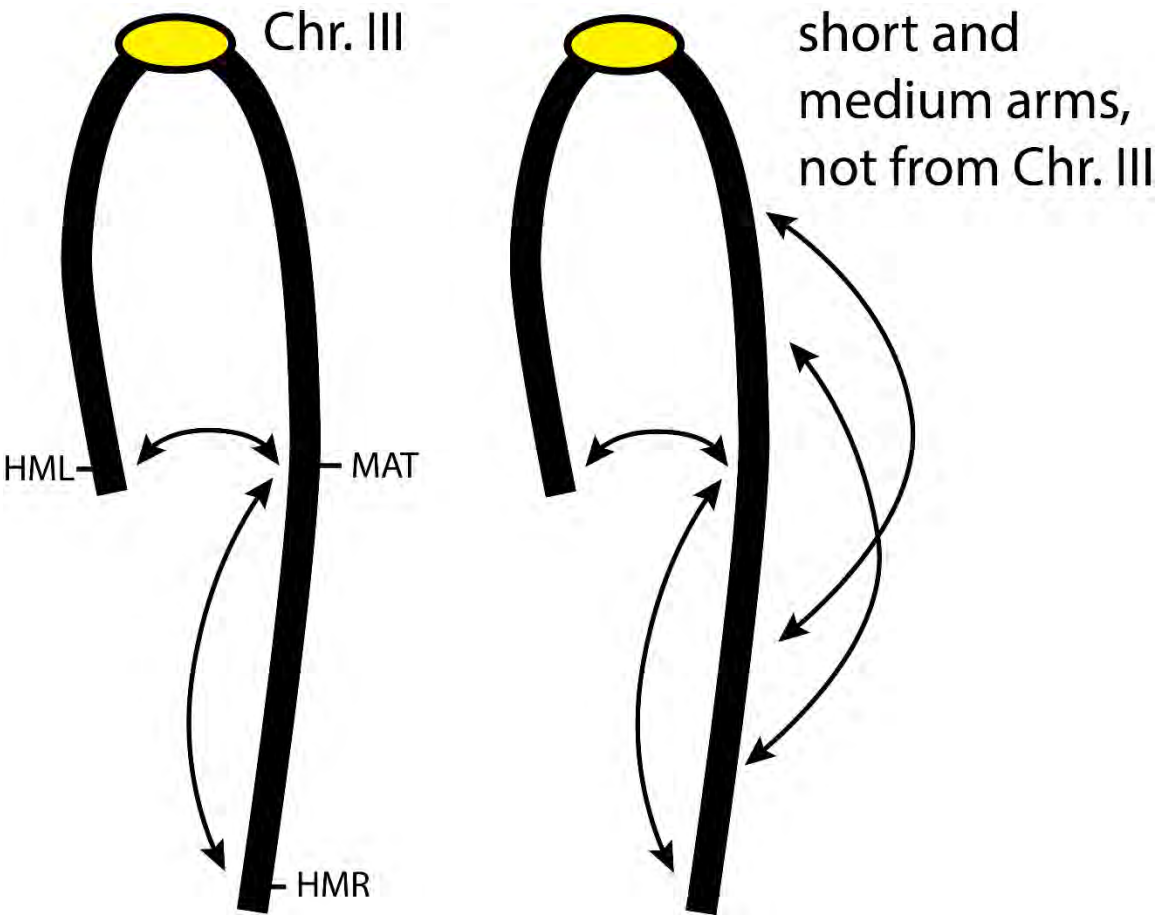


Figure 4.4. Schematic of how the expected interaction frequency for *HML-MAT* and *HMR-MAT* was calculated. Left shows a schematic of chromosome III and the black arrows indicate the contact frequencies measured between *HML* and *MAT* and between *HMR* and *MAT*. The right is a schematic of other short and medium arm chromosomes. The black arrows that are between the two arms are interaction collected from these chromosome that are at the same position from the centromere that *HML* and *MAT* are. Black arrows along a single arm indicate interaction frequencies that were collected that are the same distance that *HMR* is from *MAT*.

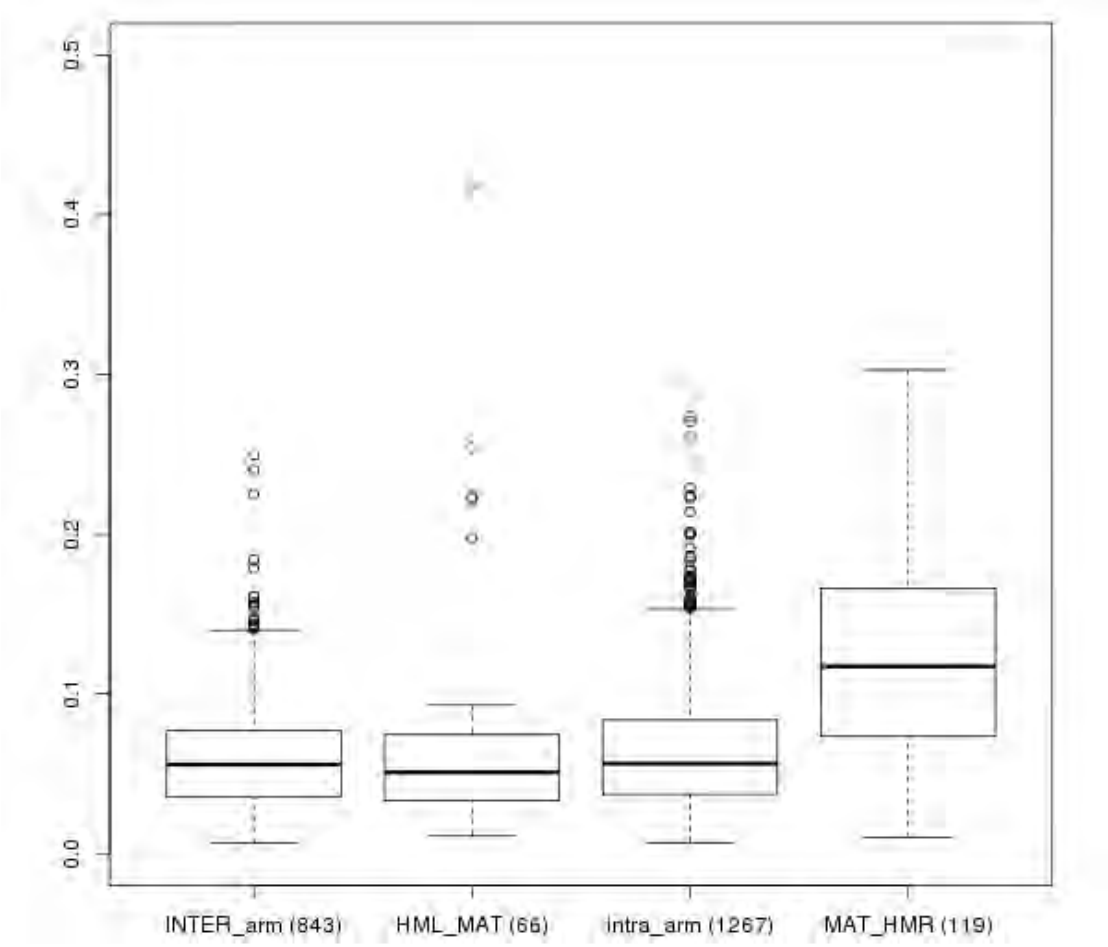


Figure 4.5. Box plots of interactions collect in the manner illustrated in Figure 4.4 in *MAT*α cells. Y-axis is the contact probability between the indicated loci. "INTER_arm" interactions are interactions between short or medium arms of the same chromosome in which both of the loci are positioned the same distance from the centromere as *HML* and *MAT* are. These interactions are the expected level of interactions between *HML* and *MAT* if there were no additional mechanism other that chromosome conformation at play. "*HML-MAT*" are the interaction on Chr. III between *HML* and *MAT*. "intra_arm" are interactions on small or medium length chromosome arms between two loci that are the same distance away that *MAT* is from *HMR*. This is the expected value for *HMR-MAT* interactions if there were not mechanism other than chromosome conformation at play. "*MAT-HMR*" are the interactions observed between *MAT* and *HMR*. The number in parentheses is the number of interactions that were collected for the analysis.

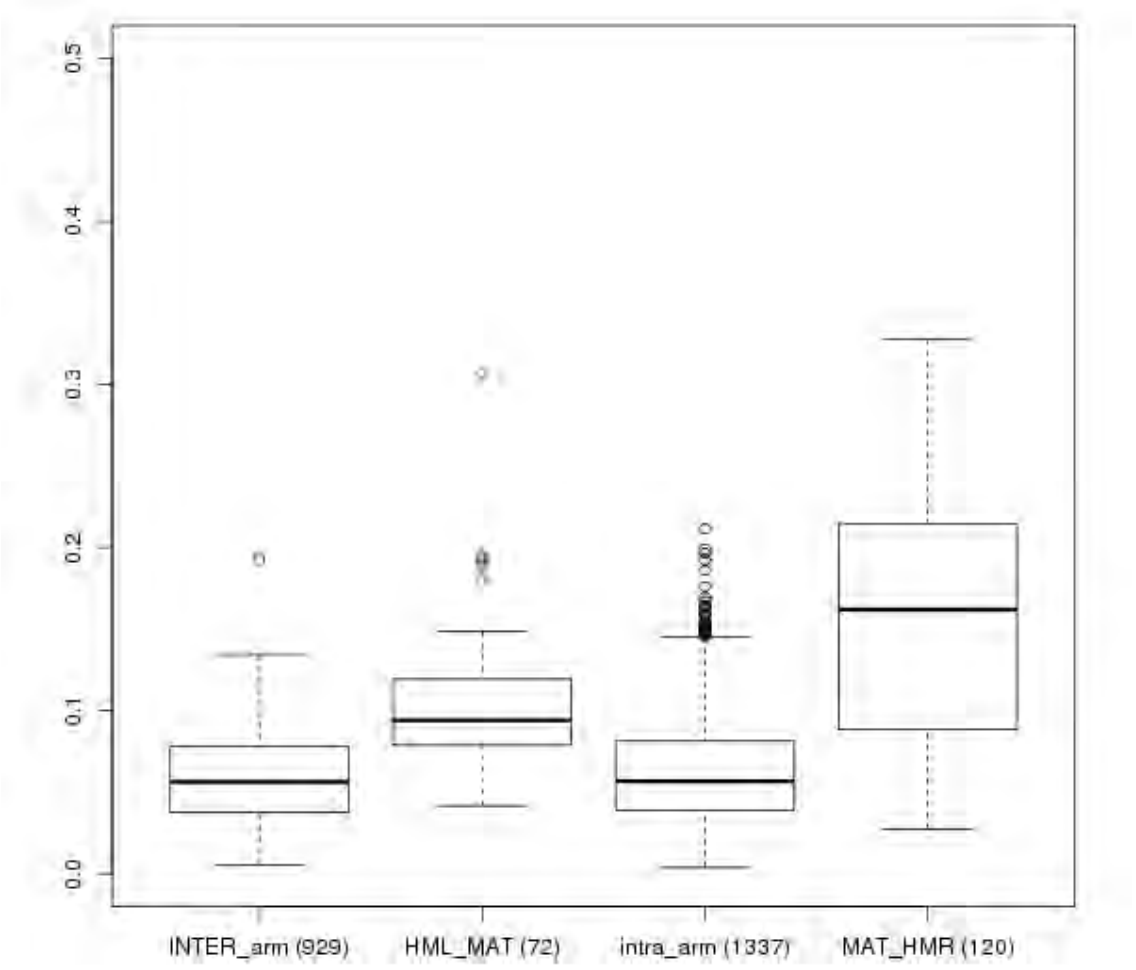


Figure 4.6. Box plots of interactions collect in the manner illustrated in Figure 4.4 in *MAT***a** cells. Y-axis is the contact probability between the indicated loci. "INTER_arm" interactions are interactions between short or medium arms of the same chromosome in which both of the loci are positioned the same distance from the centromere as *HML* and *MAT* are. These interactions are the expected level of interactions between *HML* and *MAT* if there were no additional mechanism other that chromosome conformation at play. "*HML-MAT*" are the interaction on Chr. III between *HML* and *MAT*. "intra_arm" are interactions on small or medium length chromosome arms between two loci that are the same distance away that *MAT* is from *HMR*. This is the expected value for *HMR-MAT* interactions if there were not mechanism other than chromosome conformation at play. "*MAT-HMR*" are the interactions observed between *MAT* and *HMR*. The number in parentheses is the number of interactions that were collected for the analysis.

HMR it is expected that *HML* will interact with *MAT* as often as *HML* will (**Fig. 4.5, 4.6**). This is much higher than what would be expected given the genomic distance between *HML* and *MAT*. The increase of interaction of *HML* with *MAT* is driven by the kinking of the chromosome at the centromere, which pushes the arms in towards the nuclear interior. However when we measure the *HML-MAT* and *HMR-MAT* interactions in *MAT*α cells, we see that *HMR* interacts more with *MAT* than expected and more than *HML* does (**Fig. 4.5**). Since the right arm is approximately twice the length of the left arm, it must fold back on itself in order for *HMR* to interact with *HML* and *MAT*. This folding back of the right arm, caused by the *HML-HMR* interactions, is likely what is driving the increase in interaction between *HMR* and *MAT*. This is likely the reason why *HMR* is the default donor during switching.

In *MAT***a** cells, *HML* interacts with *MAT* more than expected but not as much as *HMR* (**Fig. 4.6**). This corresponds to the major difference observed in the conformation of chromosome III between the two mating types, which is that the distal portion of the left arm, containing *HML* and the RE, interacts more frequently with an area that extends from the centromere to the *MAT* locus on the right arm (**Fig. 4.1c**). This is visible directly in the Hi-C interaction maps, and more pronounced in the Log2 difference heatmap. There is also an increase in interactions between the end of the right arm and this same region from the centromere to *MAT* (**Fig. 4.1c**). This is probably at least in part driven by the

HML-*HMR* interactions. Analysis of the similarly small chromosome VI revealed no mating type specific differences (**Fig. 4.1d**).

A more frequent association of the left arm with the centromere proximal region of the chromosome in *MAT***a** cells should also alter the interactions between this arm and arms of other chromosomes. Indeed, we find that the left arm in *MAT***a** cells interacts less with other chromosome arms than does the left arm in *MAT*α cells (**Fig. 4.2**). Thus, the left arm in *MAT***a** cells seems to be more crumpled in towards the centromere proximal region which reduces its ability to stretch out into the nuclear volume.

We sought to verify the mating type-dependent difference in chromosome conformation using an independent method. We inserted an array of 128 TetO operators just to the left of the RE (ChrIII:26,969-27,433) and an array of 256 LacO operators just to the right of the centromere (ChrIII:115,899-116,396). In this strain, we also expressed TetR-mRFP and LacI-CFP fusion proteins that bind to the corresponding operator arrays and fluorescence in live cells to indicate their position (**Fig. 4.1e**). We then measured three-dimensional distances between these two loci in both *MAT*α and *MAT***a** cells (**Fig. 4.1e**). In *MAT***a** cells, we observed a significant increase in the number of nuclei where the RE-proximal region is relatively close to the centromere (<250 nm, or 250-500 nm) as compared to *MAT*α cells, confirming our Hi-C data.

From these analyses, we conclude that chromosome III is the only chromosome that is folded in a mating type dependent manner, through differential co-localization of the left arm with the centromere proximal domain.

5C Analysis of Chromosome Organization

To further investigate the conformation of chromosome III, we performed targeted 5C analysis. We designed 5C probes to HindIII restriction fragments on chromosome III as well as for two control chromosomes V and XII (**Fig. 4.7a**). 5C provides high-resolution interaction maps in a more cost-effective manner than genome-wide Hi-C because it focuses analysis on only a select section of the genome (Dostie et al. 2006c; Dekker et al. 2013). Forward and reverse 5C probes were designed on restriction fragments in an alternating fashion (Lajoie et al. 2009b). This design prevents the detection of undigested fragments or circularized partial digestion products, which can represent a significant fraction of ligation products in 3C libraries, and provides a high-resolution and evenly spaced interaction matrix (Dostie et al. 2006c; Ferraiuolo et al. 2012b). We were able to design unique 5C probes to 569 out of 701 HindIII fragments (See Materials and Methods). Combined, these 5C probes detect 80,784 chromatin interactions that occur within and between the three selected chromosomes (**Fig. 4.1b-c**).

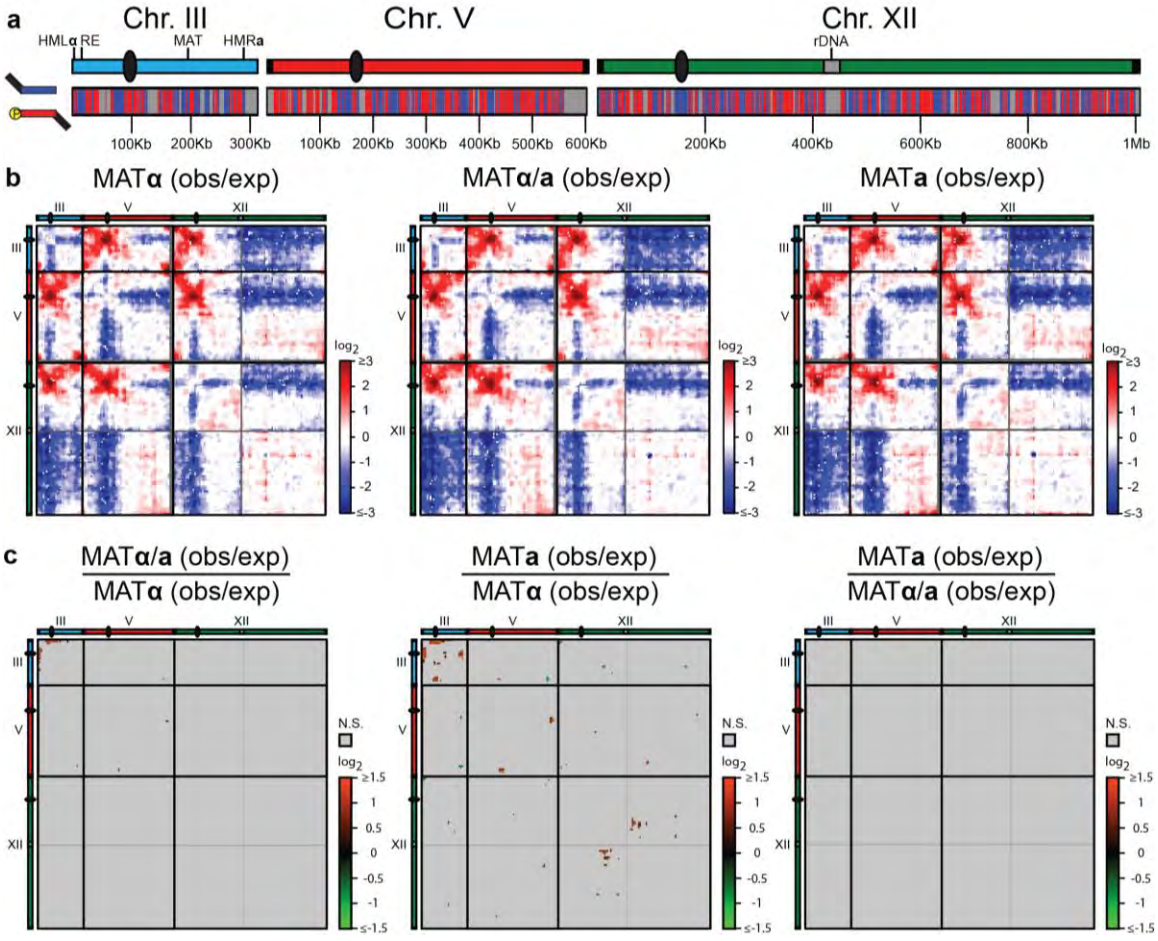


Figure 4.7. Mating-Type specific chromosome conformation in budding yeast. (**a**) 5C data normalized by the distance dependent expectation and then binned as in Figure 1. *MAT*α cells (left), *MAT***a***/α* cells (middle), and *MAT***a** cells (right). (**b**) $\log_2$ ratio of normalized 5C signal of pairs of strains. Light grey is given to bins that did not have a difference that was significant at a 5% FDR. Dark grey represents regions of no data. Bins with color are those that showed a significant difference between the two samples at a 5% FDR. The color value of significant bins is the $\log_2$ ratio of the median distance-normalized 5C signal for that bin in strain one and that of strain two. Left: comparison between *MAT***a***/α* and *MAT*α cells. Middle: comparison between *MAT***a** and *MAT*α cells. Right: comparison between *MAT***a** and *MAT***a***/α* cells.

5C interaction maps were obtained for three replicates of exponentially growing *MAT***a**, *MAT*α and diploid cells. To remove any small biases due to minor variations in 5C primer efficiency, chromatin digestion and ligation efficiency, or DNA sequencing, we used the Iterative Coverage Correction method ICE (Imakaev et al. 2012) to adjust the 5C signal in an unsupervised manner. ICE uses an iterative approach to generate balanced interaction maps where the sum of each row and column are the same. ICE operates under the assumption that each locus should be equally "visible" in a genome-wide interaction map. Previously, we successfully applied ICE to 5C maps of individual chromosomes (Naumova et al. 2013). Analysis of the Hi-C interaction maps for chromosomes III, V and XII indicates that the assumption of equal visibility is reasonably met for this subset of the genome. We applied ICE to normalize for detection biases and to balance 5C interaction maps. **Figure 4.8** shows all interaction maps for all replicates and strains analyzed in this study.

The most prominent feature in any chromatin interaction dataset is high interaction frequencies for pairs of loci that are very close to each other in the linear genome (Dekker et al. 2013). This signal decays rapidly with increasing genomic distance. This results in the strong signal along the diagonal in the interaction heatmap (**Fig. 4.9b, 4.1a, 4.8**). In order to reveal additional interaction patterns, we removed the overall distance-dependent 5C signal. This was done by first calculating the dataset-wide average relationship between genomic distance and interaction frequency (See Materials and Methods) to estimate the

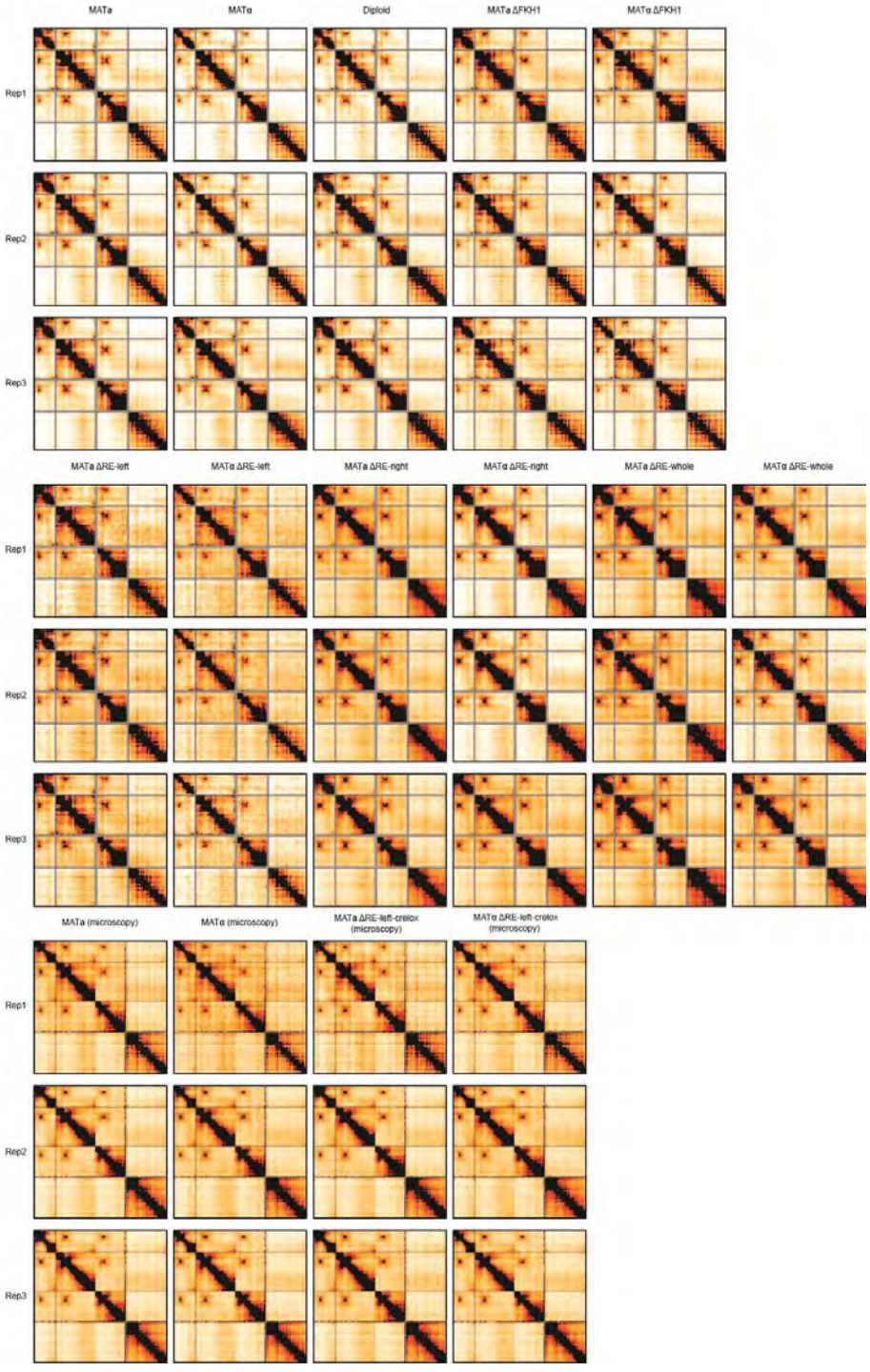


Figure 4.8. 5C Heatmaps of all datasets in this study. The data is binned at a 30kb bin that steps through the genome every 10kb. All heatmaps are normalized to the same number of reads and are drawn with the same absolute color scale so they are directly comparable to one another.

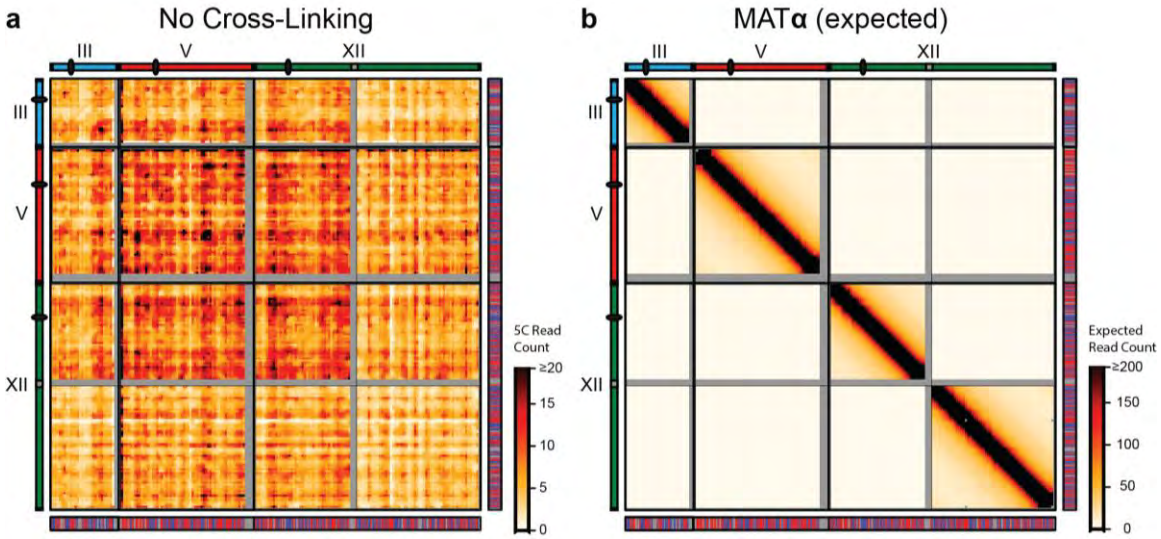


Figure 4.9. 5C controls and normalization. (**a**) One biological replicate of a 5C control library made by digesting and ligating un-crosslinked and purified genomic DNA purified from *MAT*α cells. (**b**) Heatmap of the distance dependent expectation of 5C interactions calculated from one biological replicate of *MAT*α cells. Both heatmaps are binned with a 30 kb which step over every 10kb.

expected interaction frequency of any pair of loci given their genomic separation (**Fig. 4.7b)**. The expected interaction frequency for loci located on different chromosomes was defined as the median of all observed inter-chromosomal interactions. (see Materials and Methods). We then calculated the log ratio of the observed data and the expected values for each pair of genomic loci (**Fig. 4.7b**). This analysis highlights pairs of loci that interact more or less frequently than expected given their genomic distance.

Visual inspection of the distance-corrected interaction maps for these cell types again suggests that global chromosome conformation in these strains is very similar (**Fig. 4.10**). We also confirm the depletion of interactions between the centromeres and chromosome arms and the prominent *HML-HMR* interactions on chromosome III. There is a high correlation between the 5C data and the Hi-C data (**Fig. 4.11**).

Quantification of Mating Type - Specific differences in Conformation of Chromosome III

To more rigorously analyze cell-type dependent chromatin interaction frequencies, we developed a method to identify statistically significant differences between two strains (**Methods**, **Fig. 4.12**). Briefly, we first binned the 5C data (Log_2(observed/expected values) in overlapping 30 Kb by 30 Kb bins (overlap 10 Kb, **Fig. 4.7a**), with a median coverage of ~27 pair-wise chromatin interactions

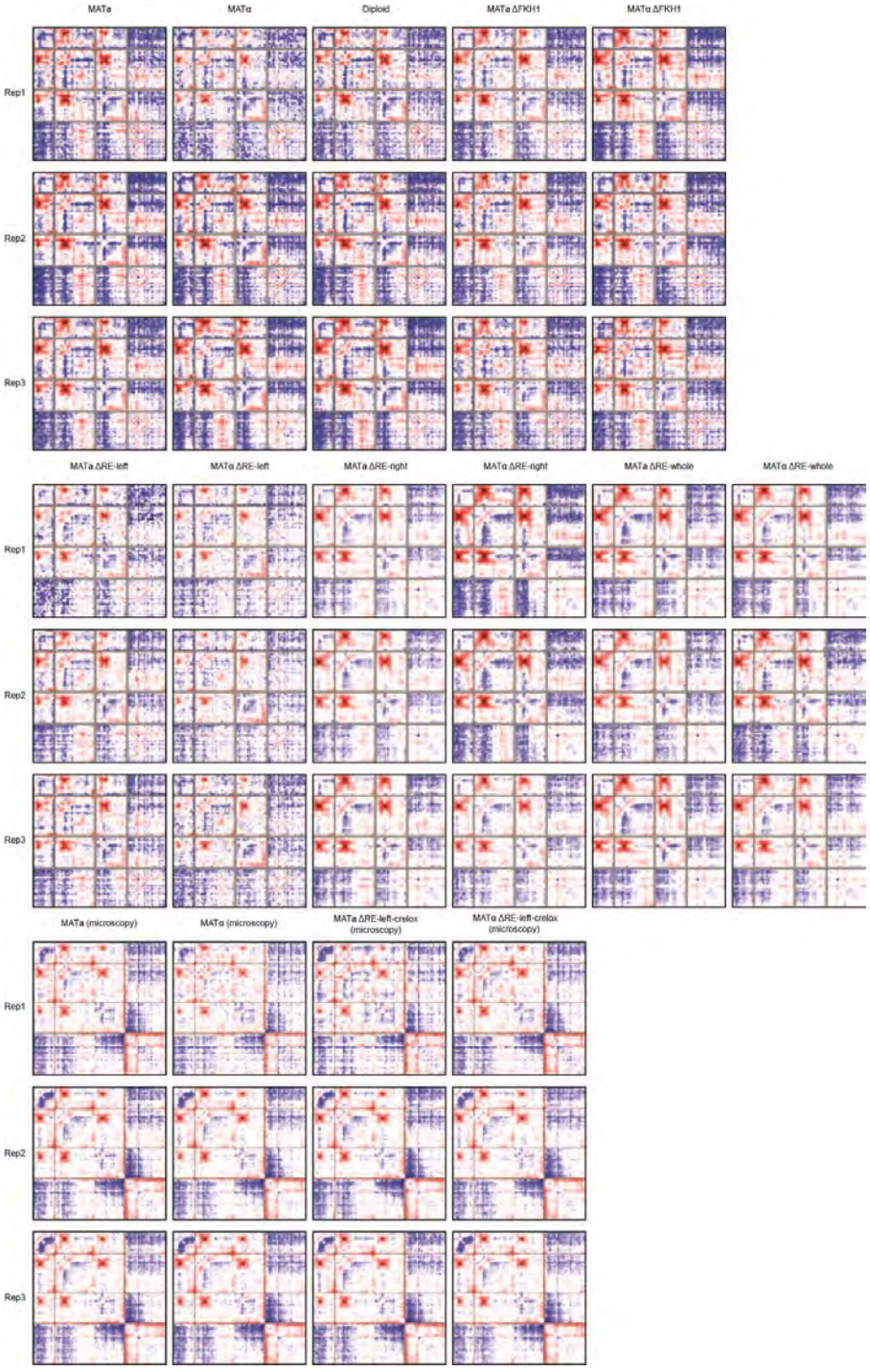


Fig. 4.10. 5C Heatmaps of all datasets in this study normalized by the expected interaction frequency at a given genomic distance. The data is binned at a 30kb bin that steps through the genome every 10kb. All heatmaps are drawn with the same absolute color scale so they are directly comparable to one another.

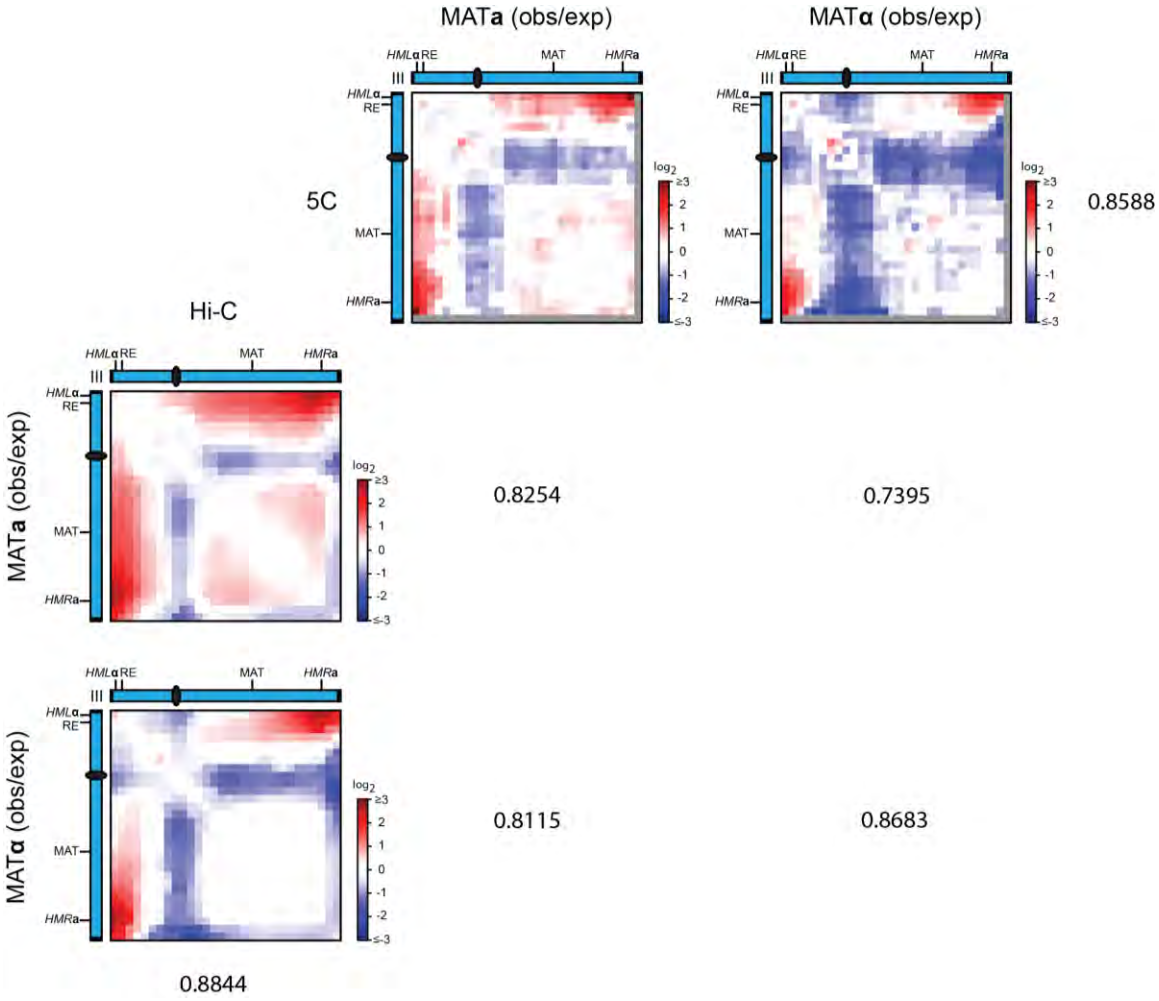


Fig. 4.11. Correlation between Hi-C and 5C data. Top shows heatmaps of *MAT***a** and *MAT*α Chr. III data normalized by the expected interaction frequency for a given genomic distance. The data is binned in a 30kb bin that steps through the genome every 10kb. Left is the Hi-C data for *MAT***a** and *MAT*α Chr. III normalized the same way as the top. The number at the intersection between a row and a column is the pearson correlation coefficient of the datasets for that row and column.

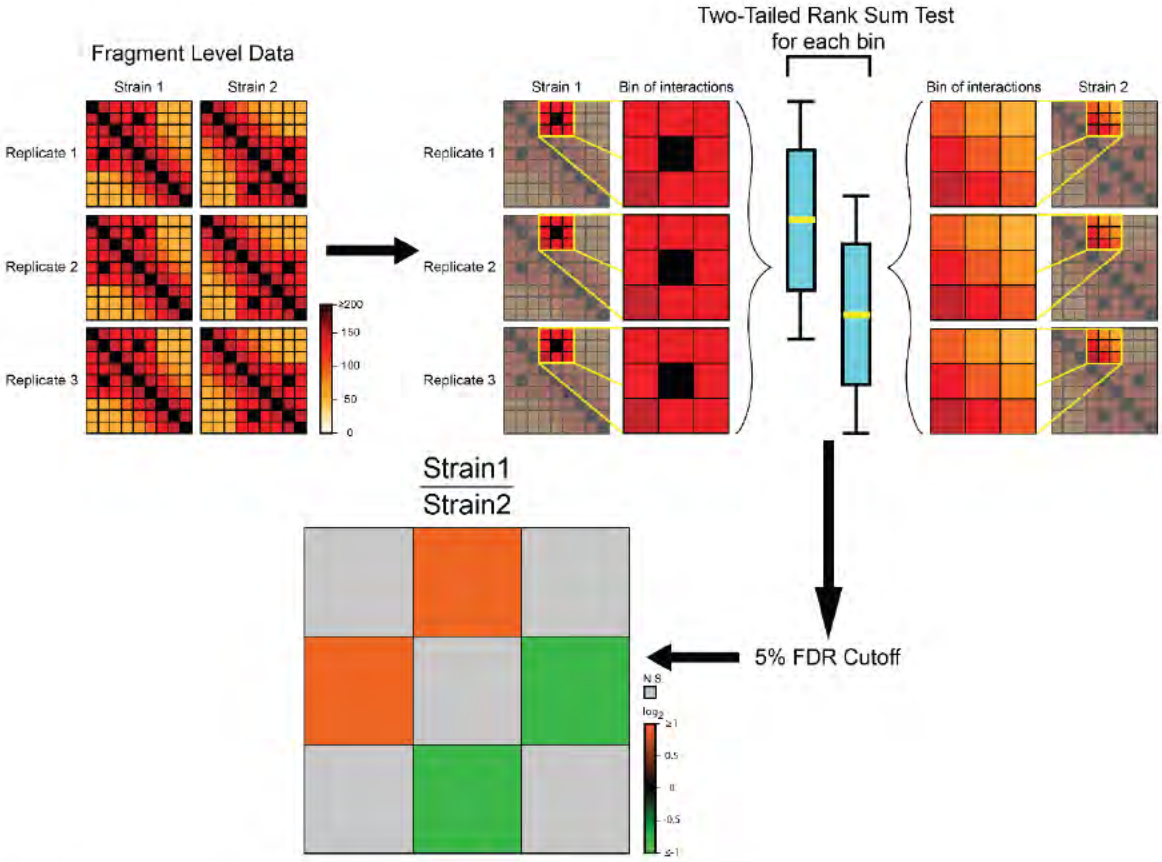
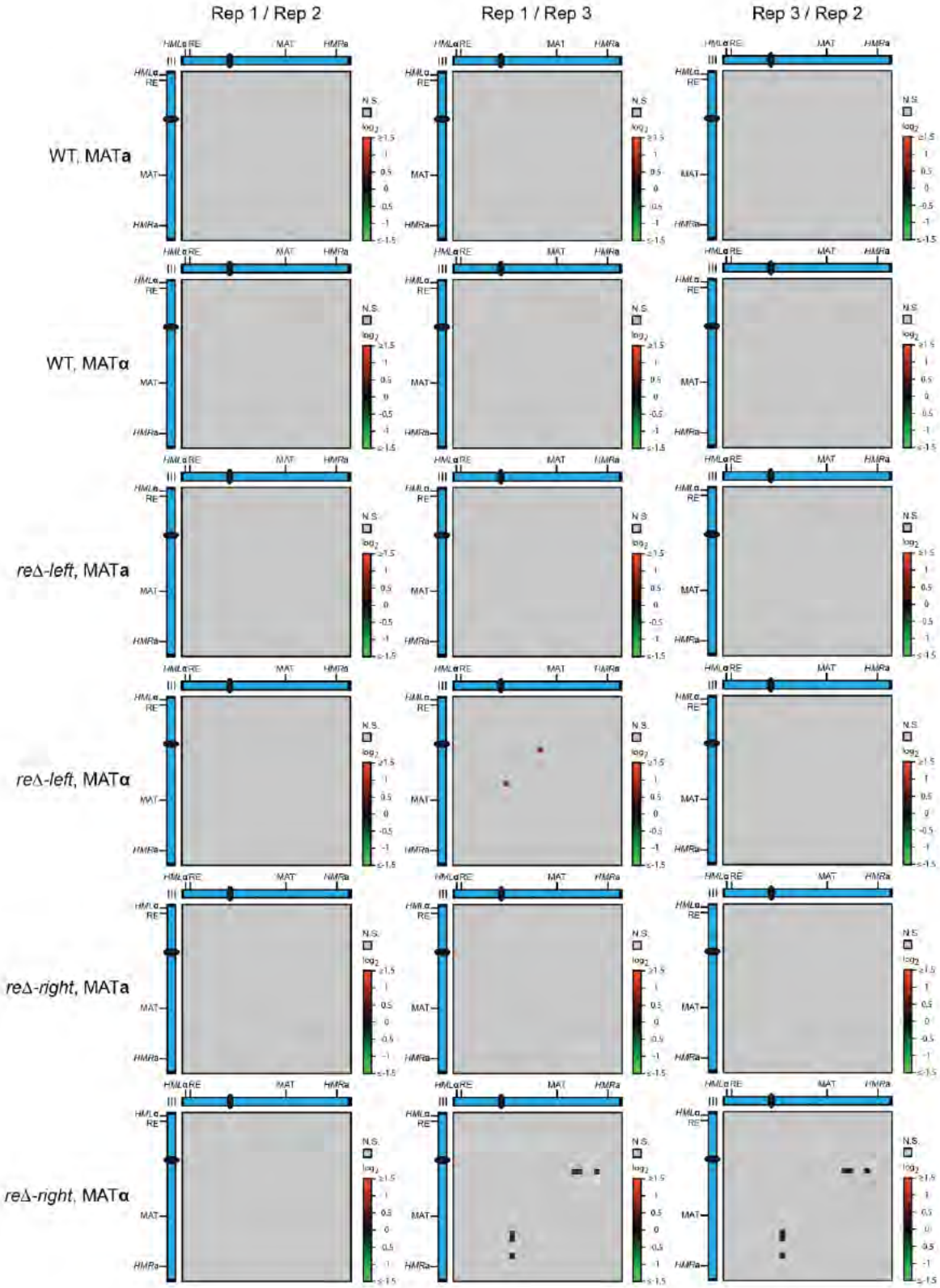


Figure 4.12. Schematic of 5C statistical analysis. Starting with fragment level data from 3 biological replicates of one strain and 3 biological replicates of another strain, the data is binned into 30Kb by 30Kb bins that step through the genome every 10Kb. Each bin of data contains all of the values from the 5C experiment from all three replicates for the given genomic region. This pooling of fragment level data for a region and across replicates increases the power of the analysis because many data points are considered for each statistical test that is done on a bin by bin basis. The median number of interactions taken from one strain for the analysis is approximately 100 per bin. A two-tailed rank sum test is used to test the difference between the two sets of data. The p-values are corrected and thresholded to a 5% False Discovery Rate (FDR). The bins for which the difference in 5C signal does not pass the 5% FDR threshold are plotted as light grey pixels in the heatmap. For those bins for which the 5C signals are significantly different the log2 of the median of strain 1 divided by the median of strain 2 is plotted in either orange for positive values or green for negative values.

pooled from all three biological replicates. We then tested whether distance-corrected 5C signals in each of the bins are significantly different between two strains at a 5% FDR threshold (**Fig. 4.12**). We then plotted only the significant differences in chromatin interactions between two strains in a heatmap where each pixel indicates the fold difference in the median 5C signal of each 30 Kb bin in strain 1 as compared to strain 2 (**Fig. 4.7b**). At this statistical threshold, very few significant differences in intra-chromosomal interactions were observed between biological replicates of the same strain indicating a low false positive rate for this test (**Fig. 4.13**).

We performed three pair-wise comparisons: Diploid vs. *MAT*α cells, *MAT***a** vs. *MAT*α cells and *MAT***a** vs. diploid cells (**Fig. 4.7**). Overall we identified a limited set of statistically significant differences between the three cell types. Two sets of cell type-specific chromatin interactions stand out. First, the regions flanking the rDNA locus interact significantly more frequently in *MAT***a** cells compared to in *MAT*α cells. This is likely due to the presence of a lower number of rDNA repeats in these *MAT***a** cells. This is not unexpected since rDNA size in *S. cerevisiae* is known to vary between populations and is influenced by growth conditions and stress from the environment (Rustchenko et al. 1993). Interaction frequencies between the rDNA flanking regions are not significantly different when comparing *MAT***a** to diploid cells or when comparing *MAT*α to diploid cells. This is likely due to the fact that the 5C signal in the diploid cells represents the average chromatin interaction frequencies from the two copies of chromosomes


Rep 1 / Rep 2
Rep 1 / Rep 3
Rep 3 / Rep 2
WT, MATa
WT, MATα
reΔ-left, MATa
reΔ-left, MATα
reΔ-right, MATa
reΔ-right, MATα
HMLα
RE
MAT
HMRa
N.S.
log2
≥1.5
1
0.5
0
-0.5
-1
≤-1.5

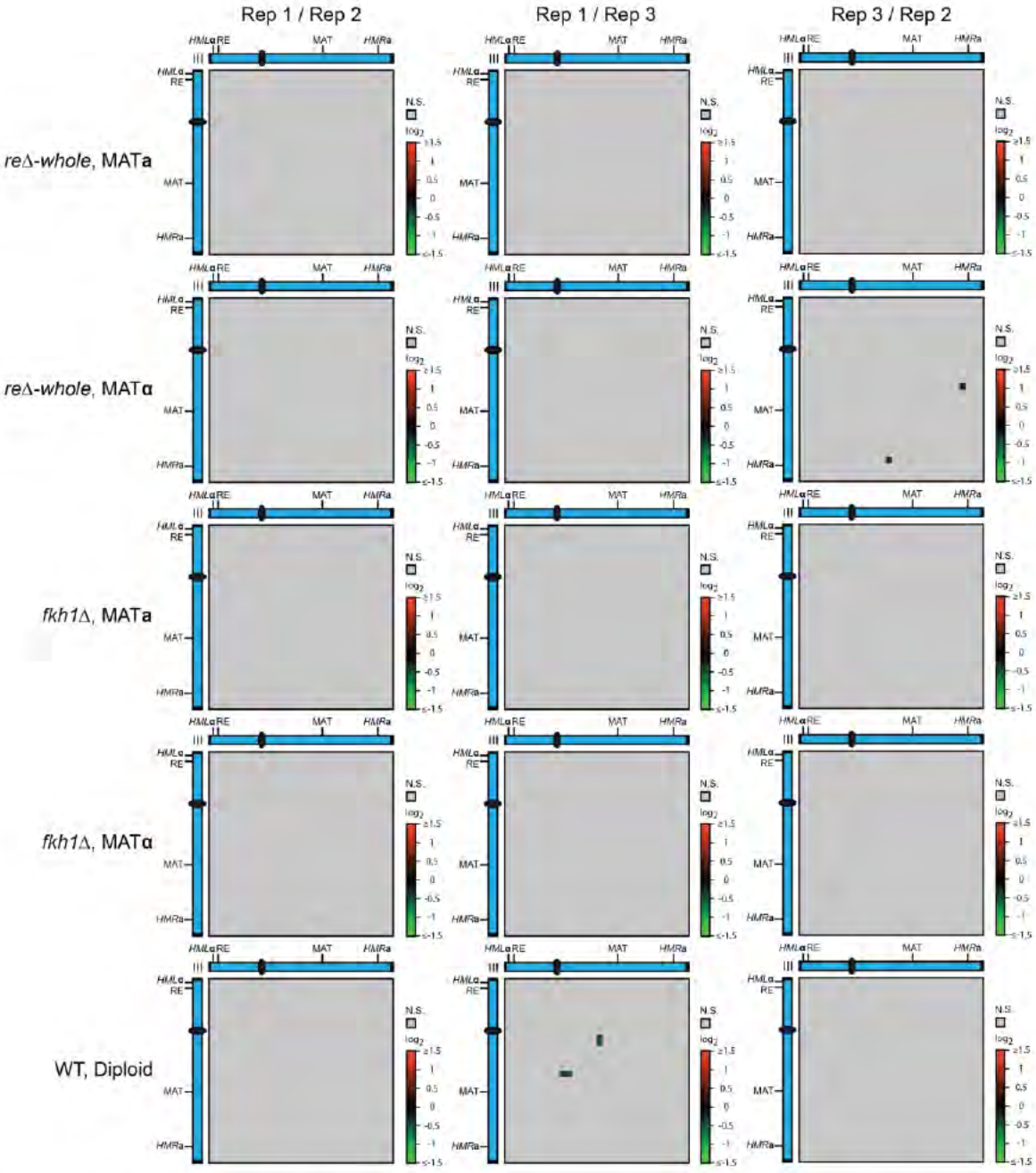


Figure 4.13: Analysis of differences between biological replicates. The same statistical method used in Figures 5.7, 5.14 - 5.19, was employed to test for differences between biological replicates. All pairwise comparisons between the replicates for a given strain are shown for chromosome III. Very few statistical differences are observed between biological replicates.

XII. Given that this diploid strain was produced by mating both the haploid strains studied here, the interactions between the chromatin segments flanking the rDNA locus are at an intermediate level compared to *MAT***a** and *MAT*α cells because the 5C interaction signal from each of the homologous chromosomes averages out. This intermediate signal is subsequently not different enough to be called significant from either haploid strain using our 5% FDR threshold.

Secondly, the conformation of chromosome III is significantly different in *MAT*α cells compared to *MAT***a** or diploid cells, with the latter being overall similar to *MAT***a** cells. In both *MAT***a** and diploid cells a portion of the left arm, including *HML* (ChrIII:0-50,000), interacts frequently with the central portion of the chromosome (ChrIII: 50,000-230,000), containing the centromere and the centromere proximal part of the right arm as well as the *MAT* locus. In *MAT*α cells these interactions are significantly lower (**Fig. 4.7, 4.14**). Thus, consistent with the Hi-C data (**Fig. 4.1**), in both *MAT***a** and diploid cells, the left arm of chromosome III is often near the centromere cluster, as also evidenced by increased interactions between the left arm of chromosome III and the centromere of chromosome III. This difference was also observed in two independent strains with a different genetic background (w303; **Fig. 4.15**). Further, and consistent with the Hi-C datasets, we again observe that in *MAT***a** cells a 50 Kb region at the telomeric end of the right arm (ChrIII:240,000-290,000) interacts with a region just around the centromere (ChrIII:70,000-150,000) (**Fig. 4.14**). Importantly, the increased interactions between the left arm

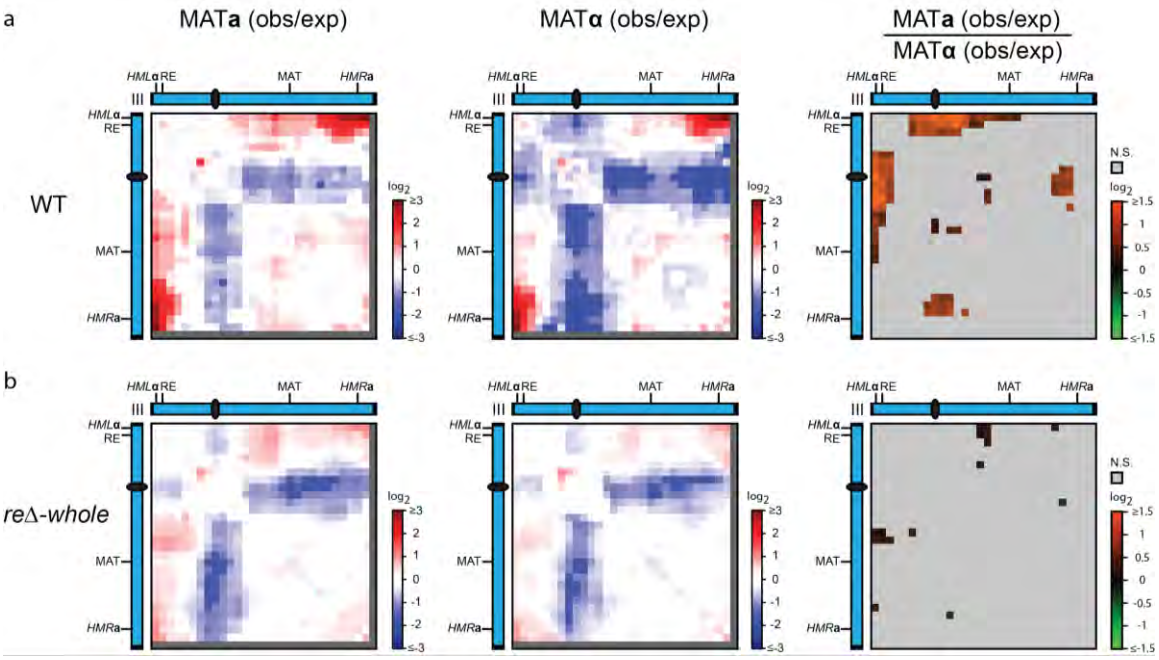


Figure 4.14. The Recombination Enhancer is responsible for mating type specific chromosome conformation. (**a**) The difference between *MAT***a** and *MAT*α conformations of Chr. III in wild type cells. (**b**) The difference between *MAT***a** and *MAT*α conformations of Chr. III in cells that have had the entire RE deleted. *Left,* is the 5C heatmap of *MAT***a** cells which has been normalized for distance dependent interaction frequency. *Middle,* is the 5C heatmap of *MAT*α cells, which has been normalized for distance dependent interaction frequency. *Right,* the log2 ratio of *MAT***a** to *MAT*α at a 5% FDR threshold.

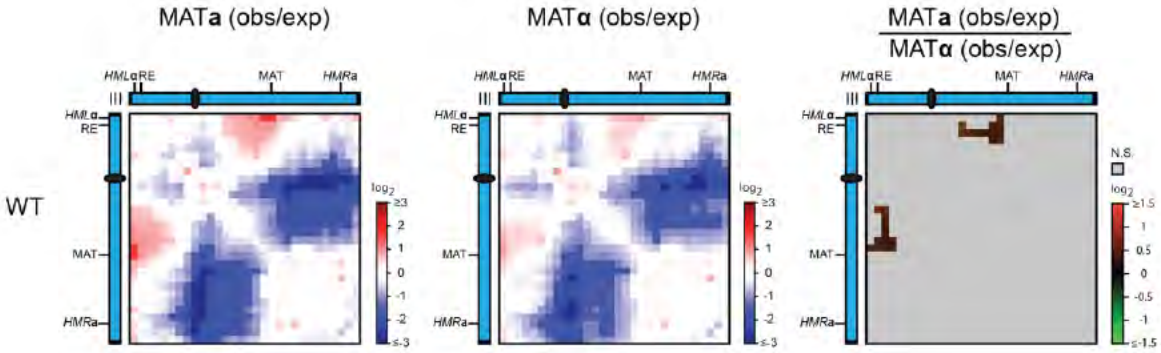


Figure 4.15. 5C analysis of strains (w303 background) used for live cell imaging shown in Figure 4.1. The observed 5C signal is plotted for *MAT***a** (left) and *MAT*α (middle). The log2 ratio of 5C signals for *MAT***a** cells divided by 5C signals for *MAT*α cells is also plotted (right). The data is binned in the same manner as in Figure 4.7. The difference in conformation of the left arm of chromosome III is also observed in this background (W303), confirming the 5C data obtained shown in Figures 5.7, which are obtained with strains with the SK1 background.

of chromosome III and the centromere proximal regions in *MAT***a** and diploid cells is not simply due to formation of a more condensed chromatin fiber as this would result in increased short-range interactions near the diagonal as well, which is not observed. Thus, the difference in chromatin interaction frequencies of the left arm reflects a change in the spatial path of the chromatin fiber itself.

The Recombination Enhancer is Responsible for the Mating-Type Dependent Conformation of Chromosome III

It has been proposed that the RE influences the conformation of chromosome III in a mating type-dependent manner so that in *MAT***a** cells the left arm becomes available for recombination during mating type switching (Haber 1998b). To test whether the RE indeed modulates the spatial organization of chromosome III we deleted the entire RE positioned between KAR4 and SPB1 (chrIII:28,988-31,140) on Chr. III (**Fig. 4.16e**) in *MAT***a** and *MAT*α cells. This deletion removes both Mcm1p/*MAT***a**lpha2p operators, DPS1 and DPS2, as well as all three Fkh1 binding site arrays. We then performed 5C on non-switching exponentially growing cultures to assess the conformation of chromosome III. Interestingly, deletion of these elements ablated the mating type specific conformation observed in wild type cells (**Fig. 4.14b**). Therefore, the RE plays a major role in modulating the mating type-dependent conformation of chromosome III. Furthermore, the conformation of the chromosome appears

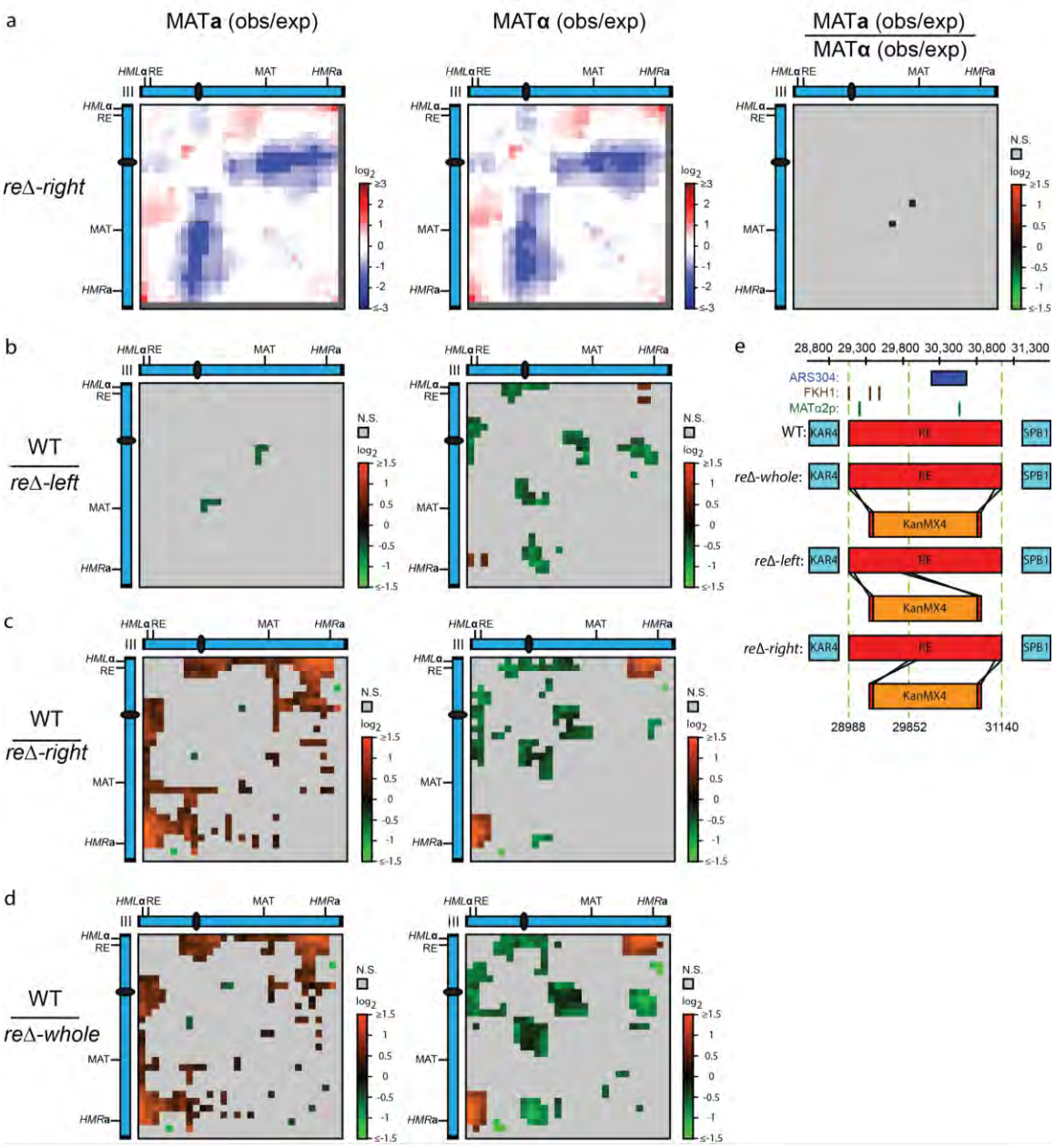


Figure 4.16. The right part of the Recombination Enhancer is responsible for the conformation of Chr. III in both mating types. (**a**) The difference between *MAT***a** and *MAT***a**lpha conformations of Chr. III in cells that have had the right part of the RE deleted. *Left,* is the 5C heatmap of *MAT***a** cells that has been normalized for distance dependent interaction frequency. *Middle,* is the 5C heatmap of *MAT*α cells that has been normalized for distance dependent interaction frequency. *Right,* the log2 ratio of *MAT***a** to *MAT*α at a 5% FDR threshold. (**b-d**) log2 ratio of the interaction frequencies between wild type cells

and mutant cells within the same mating type. Thresholded at a 5% FDR. *Left* is the comparison in *MAT***a** cells and *right* is the comparison in *MAT*α cells. (**e**) Schematic of the RE's genomic context and the particular mutants utilized in this study.

different from the wild type in both mating types (see below). The conformation of chromosomes V and XII was not affected by the deletion (**Fig. 4.8, 4.10**).

The Right Portion of the RE is the Major Driver of Mating Type-Dependent Chromosome Conformation

The 700 bp fragment on the left end of the RE, containing the Fkh1 binding sites and the DPS1, operator has been functionally defined as the minimal RE that is required and sufficient for selecting *HML* during mating type switching in *MAT***a** cells. Therefore, we hypothesized that this left portion of the RE is the driver of these mating type-dependent difference in the conformation of Chr. III that we have observed in non-switching cultures. To test this, we deleted the left portion of the RE (chrIII:28,988-29,852) in *MAT***a** and *MAT*α cells and performed 5C in exponentially growing cultures (**Fig. 4.17a**). Surprisingly, deletion of this region does not completely abolish the mating type specific conformation of Chr. III but changes were observed from the difference between the wild type *MAT***a** and *MAT*α conformations (**Fig. 4.17a**). However, all of the changes observed in this mutant are in *MAT*α cells with the *reΔ-left* deletion. Chr. III adopts a more "A-like" conformation with the left arm in closer towards the centromere proximal region and *MAT* as well as the end of the right arm comes in closer to this region also (**Fig. 4.16b**). It is important to note that the *MAT***a**-like conformation observed in the *MAT*α *reΔ-left* mutant is not a full conversion to the

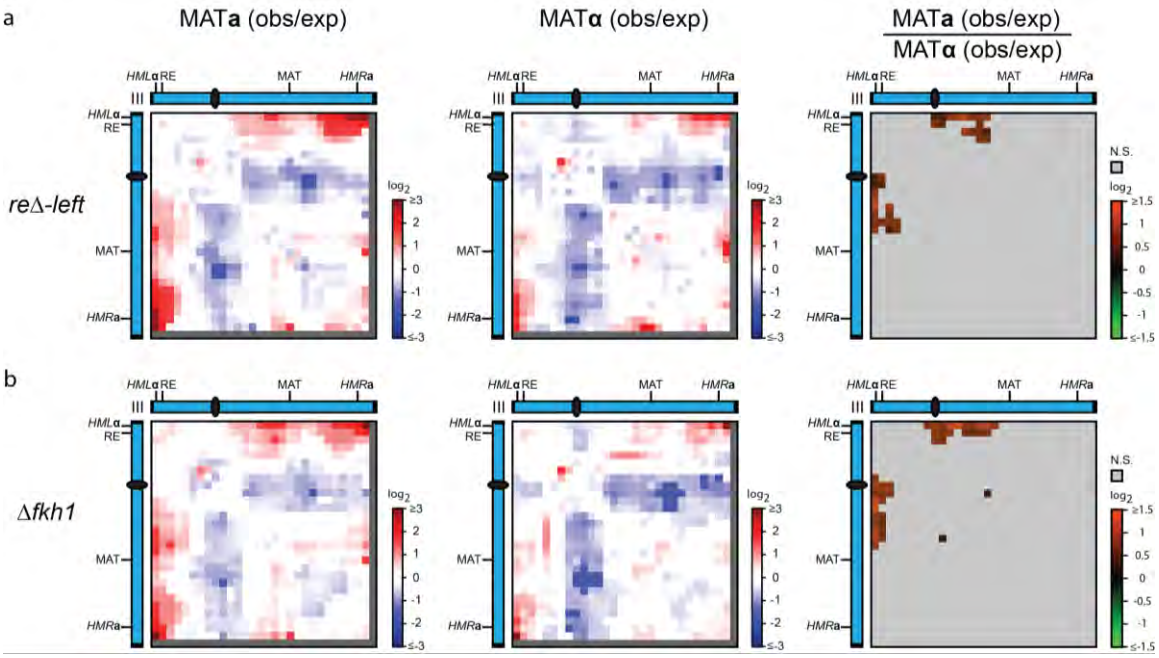


Figure 4.17. The left part of the Recombination Enhancer only plays a minor role in the conformation of Chr. III. (**a**) The difference between *MAT***a** and *MAT*α conformations of Chr. III in cells that have had the left part of the RE deleted. (**b**) The difference between *MAT***a** and *MAT*α conformations of Chr. III in cells that have had the Fkh1 protein deleted. *Left,* is the 5C heatmap of *MAT***a** cells that has been normalized for distance dependent interaction frequency. *Middle,* is the 5C heatmap of *MAT*α cells that has been normalized for distance dependent interaction frequency. *Right,* the log2 ratio of *MAT***a** to *MAT*α at a 5% FDR threshold.

*MAT***a** conformation since there are still differences observed between *MAT***a** and *MAT*α Mating types in this strain. To further study the role of the left portion of the RE in modulating chromosome conformation, we deleted the Fkh1 gene. In this mutant, the Fkh1 protein that binds the minimal RE in *MAT***a** cells is absent. 5C analysis showed that the absence of Fkh1 has a very similar phenotype to the RE-left deletion (**Fig. 4.17b**).

Since the left portion of the RE did not contribute to all of the structural differences of Chr. III between the mating types, we reasoned that the right most portion of the RE (chrIII:29,852-31,140) might involved. We deleted this region in *MAT***a** and *MAT*α cells and performed 5C on exponentially growing cultures. Deletion of this portion of the RE had the same effect as deleting the entire 2.152 Kb RE (**Fig. 4.14c, 4.16a, c**). In this mutant, the conformation of chromosome III is the same in both mating types and very similar to the conformation observed in strains that have the entire RE deleted. We conclude that the right portion of the RE is the major element that drives the mating-type specific conformation of the chromosome III.

The RE Modulates the Conformation of Chromosome III in Both Mating Types

We noted that the mating type-independent conformation of chromosome III in the *reΔ-right* deletion strains was different from the conformation observed

in either wild type *MAT***a** or *MAT*α cells. To examine this in more detail, we quantified the difference in conformation in WT and mutant cells of the same mating type (**Fig. 4.16b-d**). In *MAT***a** cells, the *reΔ-whole* mutant resulted in a statistically significant reduction in interactions between the distal portion of the left arm with the centromere-proximal domain. Conversely, deletion of the entire RE in *MAT*α cells resulted in a statistically significant increase in interactions between the left arm and the centromere proximal region as compared to WT *MAT*α cells. Given the *HML* and *HMR* interactions, the end of the right arm also interacts more strongly with a region around *MAT*. In addition, a region between the centromere and the *MAT* locus interacts more strongly with the centromere when the RE is deleted (**Fig. 4.16d**). In *MAT***a** cells, the *reΔ-right* mutation causes the conformation to lose some of its *MAT***a** characteristics (**Fig. 4.16d**) but it does not become an "α-like" conformation (**Fig. 4.18**). The same is true for *MAT*α cells, where a loss of "α" characteristics are observed (**Fig. 4.16d**) but the conformation doesn't converge to an "**a**-like" state (**Fig. 4.19**). Therefore the conformation of both *MAT***a** and *MAT*α cell in the *reΔ-right* deletion is a third and distinct conformation from either the *MAT***a** or *MAT*α conformations.

Consistent with the analysis shown in **Fig. 4.14**, deletion of the left portion of the RE had an effect only on the conformation of chromosome III in *MAT*α cells. We observed a reduced interaction between the centromere and both the left and right arms. Unexpectedly, we observed a reduction in the *HML-HMR* interactions in both *MAT***a** and *MAT*α cells when either the entire RE or the right

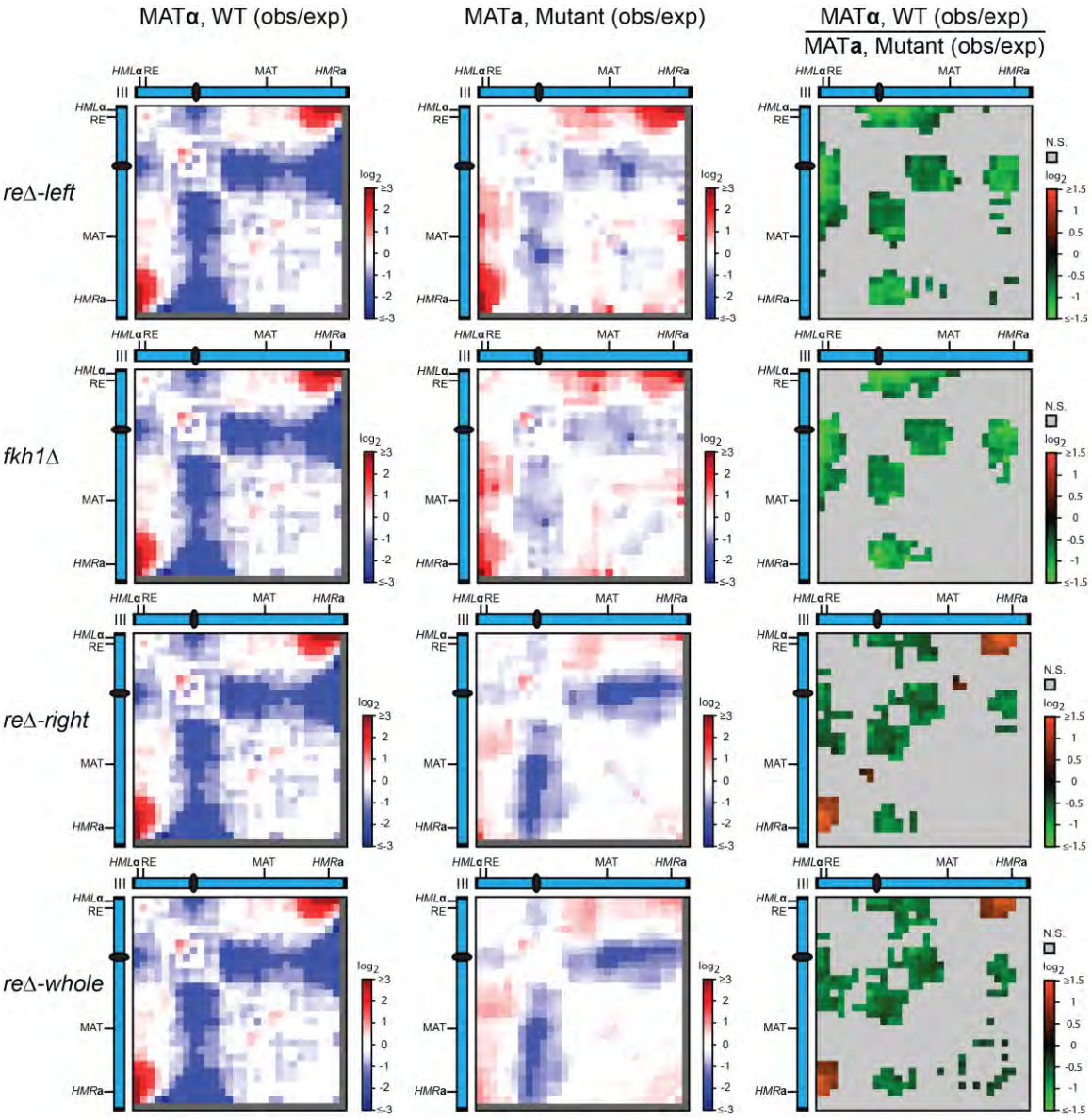


Fig. 4.18. Comparison of mutant *MAT***a** conformations to wild type *MAT*α conformations. Left column is the heatmap of Wild type *MAT*α interaction frequencies for Chr. III. The middle column is the conformation of *MAT***a** cells for the mutant located on the left. Right column is the significant differences between the Wild type *MAT*α cells and the mutant *MAT***a** cells. A completely grey heatmap for the differences would indicate that the *MAT***a** conformation in that mutant converged to the wild type *MAT*α state.

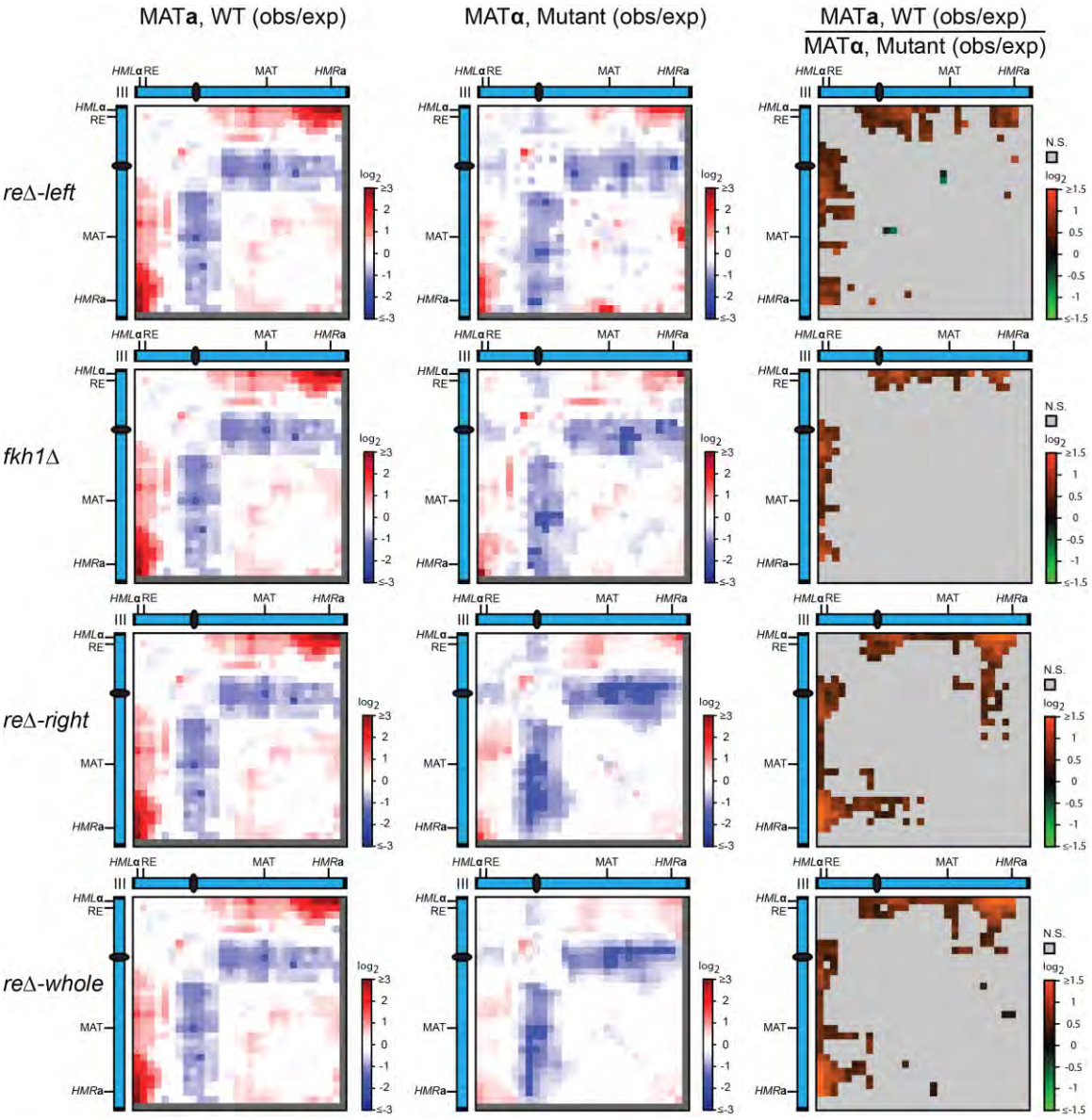


Fig. 4.19. Comparison of mutant *MAT*α conformations to wild type *MAT***a** conformations. Left column is the heatmap of Wild type *MAT***a** interaction frequencies for Chr. III. The middle column is the conformation of *MAT*α cells for the mutant located on the left. Right column is the significant differences between the Wild type *MAT***a** cells and the mutant *MAT*α cells. A completely grey heatmap for the differences would indicate that the *MAT*α conformation in that mutant converged to the wild type *MAT***a** state.

portion of the RE is deleted (**Fig. 4.14a,b and Fig. 4.16d**). Whether this is a direct or indirect effect is currently unknown.

These analyses show that the right portion of the RE puts different structural constraints on the chromosome in *MAT***a** and *MAT*α cells and that in the absence of this element the chromosome adopts a different, less constrained mating type-independent conformation.

Discussion

The fact that deletion of the right portion of the RE has the same effect as deletion of the whole RE suggest that the right portion of the RE is epistatic to the left portion of the RE.

It has long been hypothesized that the conformation of chromosome III would be mating type specific (Haber 2012b). Our comprehensive and high-resolution Hi-C and 5C studies now firmly establish that the three-dimensional organization of chromosome III, and most particularly the left arm, is dependent on the mating type of the cell. The left arm is either juxtaposed near the centromeric and surrounding central part of the chromosome in *MAT***a** cells, or is in a more extended configuration projecting away from the centromere in *MAT*α cells. We confirmed this differential positioning by live cell imaging. Imaging with three tagged fluorescent loci also provides evidence that the relative spatial distance between *HML*, *HMR* and the *MAT* locus differs in *MAT***a** cells as

compared to *MAT*α cells in non-switching cells (Lassadi et al., unpublished observations) consistent with our 5C data.

As a result of the differential folding of the chromosome, *HML* is positioned closer to *MAT* to make it better able to compete with *HMR* for interactions with *MAT*, but it doesn't allow *HML* to out compete *HMR* based on contact probability alone (**Fig. 4.6**). This indicates that the mechanism that acts to select the left arm of Chr. III for switching is functionally different from the mechanism that establishes the mating type specific conformations.

Although the left portion of the RE is essential for mating type-specific selection of the left arm for mating type switching, we find that this portion is not solely responsible for mating type dependent chromosome conformation. Deletion of the left portion of the RE did not completely abolish the differential folding of Chr. III, but did affect some aspects of its conformation in *MAT*α cells. However, deletion of the right portion of the RE alters the conformation in both mating types. Thus, our data reveal roles for the RE in modulating the conformation of chromosome III in different ways in the different mating types (**Fig. 4.21**).

Mating type switching is highly non-random so that *MAT***a** cells strongly favor recombination between the *MAT* locus and *HML* on the left arm, while *MAT*α cells favor recombination between *MAT* and *HMR* located on the right arm. This preference is position-dependent and not dependent on the DNA

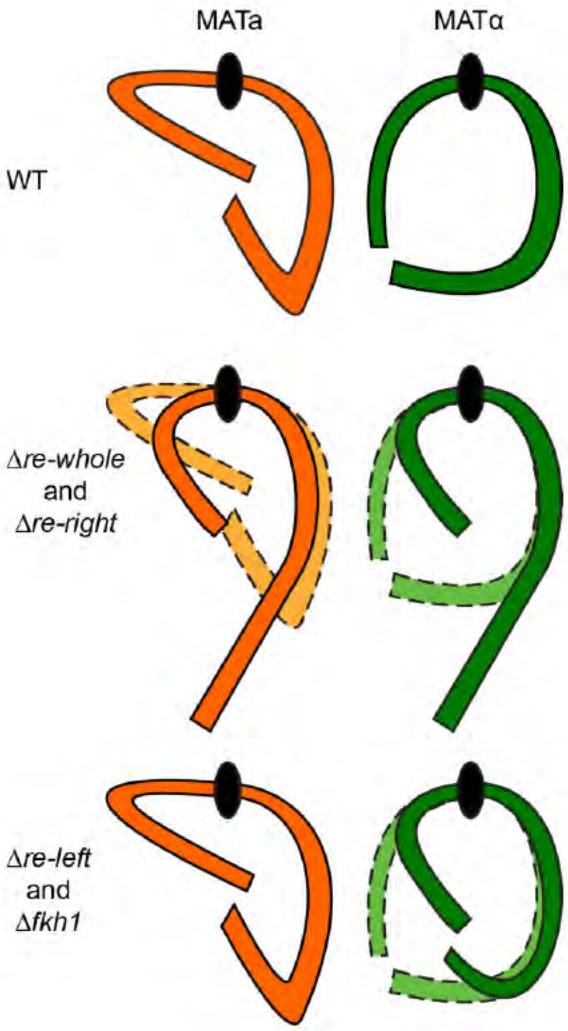


Figure 4.20. Summary of the conformation of Chr. III. *Top* schematic of the conformation in wild type cells. *Center,* schematic of the conformation in cells which have either the whole RE or the right part of the RE deleted. *Bottom* cartoon of the conformation of Chr. III in cells which have been deleted for either the left part of the RE or Fkh1. *Left* the conformation in *MAT***a** cells. *Right* the conformation in *MAT*α cells. The lighter schematics with the dotted outline are the wild type conformations for that mating type.

sequence located at *HML* or *HMR*, suggesting a role for higher order chromosome structure (Weiler and Broach 1992b). Further, it was found that the left arm is generally relatively cold for recombination, e.g. as shown by the low gene conversion frequencies between *leu* alleles placed on the left arm in *MAT*α (Wu and Haber 1995b). The left portion of the RE was found to activate the entire left arm for recombination in *MAT***a** cells (Wu and Haber 1996a). This element is both essential and sufficient for this activation. Live cell imaging experiments showed that the left arm is more mobile in *MAT***a** cells, in an RE-dependent manner (Bressan et al. 2004a). These studies suggested that the RE could act through modulating the conformation of the left arm in *MAT***a** cells.

Diploid Cells have *MAT*a Conformations of Chr. III

5C experiments of both of the haploids and the diploid strains reveal that there is a *MAT***a**/diploid conformation and a *MAT*α conformation. In *MAT***a** cells the RE is activated by the action of Mcm1p and Swi4/Swi6 (Szeto and Broach 1997; Szeto et al. 1997a; Wu et al. 1998a) and in *MAT*α cells the RE is repressed by the action of Mcm1p/*MAT*α2p/Tup1p. In diploid cells, both *MAT***a**1p and *MAT*α2p are expressed which form a complex which represses haploid specific genes. It is possible that in the diploid state *MAT***a**1p sequesters *MAT*α2p which causes a de-repression of the RE. This de-repression could then allow Fkh1 and yKu70/80 to bind to the RE which might then alter the conformation to the *MAT***a** state.

The Left Part of the RE Plays a Conformational Role in *MAT*α cells only

It was surprising that the left portion of the RE affectes the conformation of Chr. III in *MAT*α cells instead of *MAT***a** cells, which would have been predicted from switching experiments. DPS1 and DPS2 in the RE are responsible for repressing the RE. The left most site, DPS1, is a stronger repressor than the right most site and switching experiments have shown that DPS1 is essential for left arm usage in *MAT***a** cells whereas the right site, DPS2, is important for usage but not essential (Szeto et al. 1997a). One explanation of the conformational change that is observed in *MAT*α cells when the left portion of the RE is deleted is that the loss of DPS1 causes a de-repression of the RE which allows the factors responsible for the *MAT***a** conformation to bind.

Fkh1p binds to the RE in *MAT***a** cells and is essential for left arm usage during switching. Deletion of the Fkh1 binding arrays in the left portion of the RE has the same effect on left arm usage as deleting Fkh1 itself. Therefore, it is surprising that deletion of Fkh1 had the same effect on the conformation of *MAT***a** as deleting the left portion of the RE. It is puzzling how deletion of Fkh1 can cause these effects in the *MAT*α type in which it doesn't bind to the RE. One explaination for this is that Fkh1 might indirectly play a role in the conformation of Chr. III by regulating the expression of proteins that bind to the RE and modulate the conformation.

The Right Portion of the RE Modulates Chromosome Conformation in Both Mating Types.

It was interesting that deletion of the right part of the RE changed the conformation of Chr. III in both mating types. The right portion of the RE contains DPS2 as well as ARS304. The molecular mechanism that interfaces with this portion of the RE is unknown. However, ARSs play a role in other parts of the mating type switching cassette. ARSs at the HM loci recruit RAP1 which recruits Sir1 to the HM loci. Sir1 then recruits Sir2/3/4 complex which spreads and silences the HM loci (Miele et al. 2009). ARS304 might behave in a similar manner. An exploration of the factors that bind to this region will be essential to understand the mechanism that allows the RE to modulate chromosome conformation. The right portion of the RE seems to establish the conformation of Chr. III in both mating types. In *MAT*α cells, the RE causes the left arm to stretch out in such a way that it is depleted for interactions with *MAT*. The mechanism behind this is unclear and could potentially involve tethering of the left arm to a nuclear structure. Association of *HML* with the periphery is not different between the two mating types so it is unlikely that association with the periphery could cause this effect (Bystricky et al. 2009b). Also, we do not observe a significant enrichment of interactions of the left arm of the chromosome with regions adjacent to the rDNA on Chr. XII in the 5C data. In *MAT***a** cells, the RE causes the left arm to interact more with the centromere proximal region and the *MAT* locus. The fact that these interactions are over a large region and not very

strong compared to what would be expected for a point to point interaction (Doyle et al. 2014) indicates that the left arm explores more space in *MAT***a** cells than in *MAT*α cells. The mechanism of the right part of the RE that modulates chromosome conformation is independent of and redundant to the mechanism of the left part of the RE which capture DNA repairs sites to the RE. It is a very dangerous activity for a cell to intentionally make breaks in its genome that could cause deleterious rearrangements and or cell cycle arrest. Very few other biological systems, such as V(D)J recombination, intentionally make DNA breaks as part of a molecular mechanism. Chromosome conformation has been shown to facilitate recombination due to spatial proximity of genomic loci (Zhang et al. 2012a). It seems like these redundant mechanisms, that act through the RE, have evolved to ensure that the break at *MAT* is repaired as efficiently as possible. The right portion of the RE establishes mating type specific conformations that set up the chromosome spatially such that in *MAT*α cells the left arm is sequestered away from *MAT* and therefore *HMR* out competes *HML* for interactions with *MAT* and in *MAT***a** cells the left arm explores more around the *MAT* locus, which may make *HML* more competitive for interactions with *MAT*. The left portion of the RE only comes into play when *MAT***a** cells are switching. This region binds Fkh1p which may also bind to the DNA break and capture the RE to *MAT* and thus drive interaction between *HML* and *MAT*.

Materials and Methods:

Yeast cell culturing. Yeast from glycerol stocks were streaked out onto agar plates with the appropriate media and incubated at 30°C. Three colonies were picked and grown overnight in a 6ml culture of the appropriate media at 30°C while shaking at 250 RPM. The culture was backed down to an OD_{600} of 0.2 in 100ml of media and grown over night at 30°C with shaking. The culture was diluted to an OD_{600} of 0.2 in 400ml of media and incubated to mid-log phase at 30˚C with shaking.

Generation of SK1 strains used for 5C experiments.

The *reΔ-left* strains were made as follows. First, the KanMX4 gene was amplified from the pKanMX4 plasmid using these PCR primers with tails specific to the genomic sequence flanking the R: 5'-taaatccttgcttagctcaattaaatatactagtaaataacgtacgctgcaggtcgac -3' and 5'-atgtcactggacaatacaaatctttggaaaccgtccctagtcgatgaattcgagctcg -3'. This amplicon was transformed into NKY2996 (*ho::LYS2*, *lys2*, *ura3*, *nuc1Δ::LEU2*, *MAT*α). Transformants were selected on 400 µg/ml G418 plates. Deletion of the RE was tested in this strain (yAM31) by PCR using these primers: 5'-cttctctaccgcgactagacgatt -3' and 5'- ctgcagcgaggagccgtaatttt -3.

To generate the equivalent *MAT***a** strain, this strain (yAM32) was mated with NKY2997 (*ho::LYS2*, *lys2*, *ura3*, *nuc1Δ::LEU2*, *MAT***a**). The diploid strain was sporulated and tetrads were dissected and G418-resistant spores were selected.

The mating type was tested using a mating assay in order to obtain the *MAT***a** *reΔ* strain.

The *reΔ-right* strains were made as follows. First, the KanMX4 gene was amplified from the pKanMX4 plasmid using these PCR primers with tails specific to the genomic sequence flanking the region to be deleted: 5'-ATAAAATGGATATTGTTGTACTTCGGTAGATGTTATATTCTAAAGAGTTAAATTATCCAATTCCAAATTCcgtacgctgcaggtcgac -3' and 5'-ACTCAATAGTGTGAGGTTTACGTGGCTATATATACGCAATATATACACTTTAATCCAAATTATCCCTCCCtcgatgaattcgagctcg -3'. This amplicon was transformed into NKY2996 (*ho::LYS2*, *lys2*, *ura3*, *nuc1Δ::LEU2*, *MAT*α) and NKY2997 (*ho::LYS2*, *lys2*, *ura3*, *nuc1Δ::LEU2*, *MAT***a**). Transformants were selected on 200 µg/ml G418 plates. Deletion of the RE was tested in this strain by PCR using primer pairs: 5'- CGTGTTCTAGCGTGTTCACT -3' and 5'-CAATTCAACGCGTCTGTGAG -3' and 5'- GAATGCTGGTCGCTATACTG -3' and 5'- GGATGAACGAGCACGATAAC -3'.

The *reΔ-whole* strains were made as follows. First, the KanMX4 gene was amplified from the pKanMX4 plasmid using these PCR primers with tails specific to the genomic sequence flanking the region to be deleted: 5'-GTACAAAAAGCACAATAAAGCATCAACACATAAATCCTTGCTTAGCTCAATTAAATATACTAGTAAATAAcgtacgctgcaggtcgac -3' and 5'-

ACTCAATAGTGTGAGGTTTACGTGGCTATATATACGCAATATATACACTTTAA TCCAAATTATCCCTCCCtcgatgaattcgagctcg -3'. This amplicon was transformed into NKY2996 (*ho::LYS2*, *lys2*, *ura3*, *nuc1Δ::LEU2*, *MAT*α) and NKY2997 (*ho::LYS2*, *lys2*, *ura3*, *nuc1Δ::LEU2*, *MAT***a**). Transformants were selected on 200 µg/ml G418 plates. Deletion of the RE was tested in this strain by PCR using primer pairs: 5'- ACCGCGACTAGACGATTACA -3' and 5'- CAATTCAACGCGTCTGTGAG -3' and 5'- GAATGCTGGTCGCTATACTG -3' and 5'- GGATGAACGAGCACGATAAC -3'.

The *fkh1Δ* strain was made with a two-step PCR stitching method. First the primers: 5'- TTTGGTAGCACAGAGGGTACAGAAGT-3' and 5'- AGGTGCGACAATCTATCGATTGTATG-3' were used to amplify the left portion of the of the *fkh1*::KanMX locus from the *fkh1Δ* strain from the yeast deletion collection. Another primer pair: 5'- AGTGACGACTGAATCCGGTGAGAATG-3' and 5'-AAACGGTATAGAGAGAACAGGATGGT-3', was used to amplify the right part of the *fkh1Δ* locus from the yeast deletion collection. Since these amplicons overlap one another, they were used as a template in a PCR reaction without any additional primers. This reaction was run for 10 cycles of denaturation, annealing, and extension, which favored the production of the full length amplicon. Then the primers: 5'- TTTGGTAGCACAGAGGGTACAGAAGT-3' and 5'-AAACGGTATAGAGAGAACAGGATGGT-3' were added to the reaction and PCR amplified for 30 cycles to produce a full length amplicon that contains the

KanMX gene with ~100bp of flanking genomic DNA on both sides from the regions just outside of the *Fkh1* locus. This amplicon was used to transform both *MAT***a** (NKY2997) and *MAT*α (NKY2996) cells. Transformants were selected on G418 plates. Primers: 5'-CTGTAGCTTTCGCGTTTGGT-3' and 5'-AGGTGCGACAATCTATCGATTGTATG-3' were used to verify the insertion of KanMX into the *Fkh1* locus in the resulting clones.

For all of the SK1 strains used in the 5C experiments, a plasmid was transformed into the cells. This plasmid has the *HO* gene under the control of the *GAL4* promoter. This plasmid also has the URA3 gene and is selected for by growing the cells in –ura media with 2% Dextrose, which actively represses the GAL4 promoter.

Mating Assay to Verify the Mating Type of Each Strain. All strains were replica plated on to lawns of the tester strains: NKY101 (*MAT***a,** *ade8*) and NKY102 (*MAT*α**,** *ade8*) on minimal media. Strains that have the pGal:HO plasmid were first plated on media containing 5FOA to select for cells that have lost this plasmid.

Genotyping of 3C Templates. All 3C templates were verified to have the correct genotype using PCR. Primers 5'- cttctctaccgcgactagacgatt -3' and 5'-ctgcagcgaggagccgtaatttt -3' were used to verify the deletion of the

Recombination Enhancer and insertion of the KanMX gene. Primers: 5'-GCATAGTCGGGTTTTTCTTTT -3' and 5'- ATCCGTCCCGTATAGCCAAT -3' were used to verify the presence of the *MAT***a** allele and Primers: 5'-TGTCTTCTCTGCTCGCTGAA -3' and 5'- ATCCGTCCCGTATAGCCAAT -3' were used to verify the presence of the *MAT*α allele. Primers: 5'-CAATTCAACGCGTCTGTGAG -3' and 5'- TCCATGGATGAGGTAAACAAAA -3' were used to detect the deletion of Fkh1 by the insertion of KanMX. As an internal control the primers: 5'- GCCTTCTAGGTTTCCATCCA -3' and 5'-GGCCAAATCGATTCTCAAAA -3' were used to detect the ACT1 gene.

Hi-C.

The yeast culture was crosslinked with 37% formaldehyde to a final concentration of 3% in the culture media for 20 min at 25˚C. The crosslinking was quenched by adding 2X the volume of formaldehyde used of 2.5M Glycine to the culture and incubating the culture for 5 min at 25˚C. Cells were lysed by grinding with a cold mortar and pestle. The chromatin was solubilized by adding 1% w/v SDS and incubating at 65˚C in a water bath for 10 min. Excess SDS was quenched using 10% v/v Triton X-100. The chromatin was digested with HindIII at 37˚C overnight in a water bath. Digested ends were filled in with biotin-14-dCTP with Klenow at 37˚C for 2 hr in a water bath. The klenow was inactivated by incubating in a 65˚C water bath for 20 min. Crosslinked and digested fragments were ligated together in a dilute reaction in a 16˚C water bath for 8 hr in order to favor intra-molecular ligation. Crosslinks were reversed by incubating

the ligation mixture with proteinase K in a 65˚C water bath. Ligation products were purified with phenol (pH 8.0):chloroform and precipitated with sodium acetate pH 5.2 and ethanol. The precipitated DNA was resuspended in 1X TE and then the DNA was re-extracted with phenol (pH 8.0):chloroform and precipitated with sodium acetate pH 5.2 and ethanol. The DNA pellet was dried thoroughly and excess salt was washed out of the solution using an Amicon 30kDa spin column using 1X TE. Then the sample was concentrated using a 30kDa Amicon column. RNA was degraded using DNase free RNase A at 37˚C for 1 hr. The Hi-C library was quantified on a 0.8% agarose gel in 0.5X TBE by comparing it to known concentration standards. The Hi-C efficiency was accessed by first PCR amplifying a neighboring interaction in the Hi-C library. This amplicon was split and the aliquots were digested with either HindIII, NheI, or HindIII and NheI at 37˚C overnight. The digested products were quantified on a gel and the percent Hi-C efficiency was calculated as the percent digested with NheI divided by the percent digested in the NheI and HindIII combined reaction. Biotin was removed from un-ligated ends using T4 DNA polymerase at 20˚C for 4hrs and then at 75˚C for 20 min to inactivate the enzyme. The DNA was sheared to 50-700bp using the Covaris S2 sonicator. The DNA ends were repaired using T4 DNA Polymerase, T4 Polynucleotide Kinase, and Klenow fragment of DNA Polymerase I at 20˚C for 30 min. The DNA was purified with one Qiagen MinElute Column. A dATP was added to the 3’ end of the molecules by using Klenow Fragment (exo-) at 37˚C for 30 min and then incubating at 65˚C

for 20 min to inactivate the enzyme. The end repaired and A-tailed library was fractionated to 100-300bp using 0.9x and then 1.1x Ampure XP extractions. For each 1.0 ng of biotinylated ligation products, 1.0 µl of Dynabeads MyOne Strepavidin C1 beads was used to enrich the library. Illumina PE adapters were ligated to the library using NEB Quick Ligase at room temperature for 15 min. The library was amplified using Illumina primers PE1.0 and PE2.0 for as few cycles as possible. Cycle number was determined by a titration experiment that revealed which cycle would produce enough library for sequencing but didn't produce higher molecular weight artifacts. The library was purified once more with 1.8x Ampure XP to remove primers and primer dimers. The final library was quantified on a bioanalyzer. The libraries were sequenced on either the GAII or HiSeq platforms. The Hi-C data was corrected using the Iterative Correction Algorithm (Imakaev et al. 2012). Brief, the data was binned into 10kb x 10kb regions and the sum of all of the interactions in each 10kb x 10kb region was used to represent the bin. The sum of all rows and columns was normalized to the average of row sum by multiplying a factor specific to each row which will raise or lower it to the average of the row sums. This process was iterated until there was very little deviation in the row sums.

Chromosome Conformation Capture (3C). The yeast culture was crosslinked with a final volume of 3% formaldehyde (37% w/v, Fisher, Catalog #: BP531-500) in the culture media at 25˚C for 20mins with shaking. The remaining formaldehyde was quenched with an excess of 2.5M glycine. The yeast cells

were harvested by centrifugation and lysed by grinding with a mortar and pestle in the presence of liquid nitrogen. The chromatin was digested with HindIII (NEB, R0104S) and the resulting crosslinked restriction fragments were ligated with T4 DNA polymerase (Invitrogen, 15224-090) in a dilute reaction so that intra-molecular ligation was favored. The crosslinking was reversed by incubating overnight at 65˚C in the presence of proteinase K (Invitrogen, 25530-015). The chimeric DNA molecules were purified with a series of phenol:chloroform pH 8.0 (1:1) extractions and precipitated with 100% cold ethanol and 3M sodium acetate pH 5.2. The RNA was degraded with RNase A (Roche, Catalog #: 10109169001) and the yield was quantified using molecular weight standards run on an agarose gel. 3C-PCR was performed on all 3C libraries with 3C PCR primers that represent a variety of genomic distances. Libraries that showed the characteristic, exponential decay of 3C-PCR product abundance relative to genomic distance passed quality control.

5C probe design. 5C probes were designed for chromosomes III, V, and XII in an alternating scheme using the S288C genome sequence (sacCer1 assembly) using the My5C design and analysis suite(Lajoie et al. 2009a). Probes were designed using the default settings of my5C: U-BLAST, 3; S-BLAST, 50; 15-MER, 800; MIN_FSIZE, 1000; MAX_FSIZE, 15,000; OPT_TM, 65; OPT_PSIZE, 40. The 5C probes contain up to 40 bases that are specific for the corresponding restriction fragment. If a shorter sequence was sufficient to obtain a predicted

annealing temperature of 65 °C, that shorter sequence was used, and random sequence was added to make a total of 40 bases. A 6 bp barcode was added to increase mappability, and a universal sequence to allow for the probes to be amplified using universal primers. The design is alternating which means that, for the most part, every other probe in the genome is a forward probe and the intervening probes are reverse probes. A forward probe means that the universal tail and barcode are on the 5' end of the molecule and the annealing sequence is complementary to 3' end of the Crick strand of its corresponding restriction fragment. A reverse probe means that the universal tail and bar code are on the 3' end of the molecule and the annealing sequence is complementary to the Watson strand of the 3' end of its corresponding restriction fragment. Fragments that were smaller than 53 bp and larger then 18,045bp were not used to design primers in order to mitigate any bias due to fragment size. Also, 5C probes were designed to have a 45-55% percent GC and to blast uniquely to the sacCer1 genome.

Chromosome Conformation Capture Carbon Copy (5C). For one 5C reaction, 4 million genome copies (53.3ng) worth of 3C library were mixed with 1fmol of each of the 5C probes in a 40 µl reaction. The reaction was heated to 95˚C for 9mins to melt the DNA, then slowly cooled to 55˚C to allow for the 5C probes to anneal to the 3C ligation products for a minimum of 4hrs and a maximum of 16hrs. The annealed probes were then ligated together using Taq DNA Ligase (New England Biolabs, M0208S) at 55˚C for 1hr. The 5C products

were amplified with a high fidelity polymerase (Amplitaq Gold, Life technologies, 4398813) and 1.6μM of each of the universal primers (Universal_forward: 5' - /5Phos/CCTCTCTATGGGCAGTCGGTGAT – 3', Universal_reverse: 5' – /5Phos/CTGCCCCGGGTTCCTCATTCTCT – 3'). Five 5C reactions for a total of 20 million genome copies were used to produce each 5C library. The Illumina Paired-end sequencing adapters and primers were used to add the Illumina Paired-end sequences to the 5C molecules in order to prepare them for sequencing on either the Illumina GAII or HiSeq platforms.

Mapping of sequence reads to 5C probe pool.

5C data was sequenced and mapped as previously described (Sanyal et al. 2012).

Briefly, fastq files were mapped to the 5C probe pool using our in-house 5C mapping pipeline. Each side of the paired-end reads was mapped independently to the 5C probe pool using the novoalign mapping software (V2.05 novocraft.com) and 5C interactions were assembled from those paired end reads where both sides mapped uniquely to a single 5C primer of opposite type (forward-reverse or reverse-forward).

Lifting over of 5C probes to the SK1 and W303 genome. Since our experiments were carried out using strains that have a different background than that for which the probes were designed to, we lifted the probe coordinates over to the correct genome assembly. For SK1 the 5C probes were aligned to the SK1 genome produced by Scott Keeney's Group (http://cbio.mskcc.org/public/SK1_MvO/) using the BLAST command line tool *blastall* (http://www.ncbi.nlm.nih.gov/staff/tao/URLAPI/blastall/) using the *blastn* program and a 20bp word size. This genome assembly contained what appeared to be a 29,701bp translocation from chromosome XI inserted at position 3,623 in the left arm of chromosome III. Our 5C data did not show evidence for a large insertion in chromosome III. Therefore this region of the assembly is likely an assembly error so we manually removed this region from the SK1 genomic sequence (chrIII:3,623-33,324). .

Once the probes were aligned to the genome, they were manually curated to remove those that blasted to the wrong chromosome or to multiple chromosomes. Probes were also filtered if they no longer aligned adjacent to a HindIII restriction site in the SK1 genome. Finally, since the length of the rDNA on the right arm of chromosome XII contains many copies of the rDNA repeats that could, in reality be up to 2Mbps of DNA, we decided to treat the left and right portions of chromosome XII as separate chromosomes in our analysis. We did this by changing the chromosome name of these probes to the right of the rDNA to chr12R and restarting the coordinates at 0.

The procedure for mapping to the 5C datasets from the strains used for the microscopy experiments (Figure 4.15) was the same as for mapping to the SK1 genome, except that the W303 genome was used (Lang et al. 2013).

Filtering, correction, and normalization of 5C data.

Prior to correcting the 5C counts using the Iterative Correction algorithm it was necessary to filter the data of any outliers. First, the 70kb around Centromere-centromere and telomere-telomere interactions were removed to prevent from filtering high 5C counts from these regions that are known to interact and thus will have a higher signal than expected. Then single data points that were greater than or equal to 7 standard deviations from the local mean, based on genomic distance, were removed for intra-chromosomal interactions. For inter-chromosomal interactions, a cutoff of 8 standard deviations from the mean of the inter-quartile (25%-50%) were removed. Finally, we removed the highest and lowest 7.5% (15% total) of rows and columns from the dataset that were determined by summing all of the inter-chromosomal interactions for each 5C fragment. The centromeres and telomeres were replaced and used for all further correction/analysis.

The data was binned into 10kb x 10kb bins. The median of the interactions in each 10kb x 10kb region was used to represent that region. The

iterative correction algorithm assumes that all bins should have the same sum of interactions. This assumption is valid when genome wide Hi-C data is being considered. However, the 5C design is comprehensive enough to approximate this assumption. The row and column sums were calculated for each 10kb bin. Then each row and column was multiplied by a factor that would raise or lower its sum to match the average of all the sums. This was done iteratively until the row and column sums deviated very little from each other.

The 5C data for each dataset was first normalized to a standard read depth. Each interaction in the dataset was calculated as a percentage of the total number of reads for the dataset. This percentage was then multiplied by the arbitrary value of 1,000,000.

Finally, 5C data was normalized relative to the expected value for pairs of loci separated by a give genomic distance (as decribed in (Sanyal et al. 2012)). This distance-dependent expected value was calculated using the “robust lowess” smoothing method which takes an ‘outlier filtered’ weighted average across a percentage of the data (here we used 5%) for each genomic distance.

5C data analysis. In order to call significant differences between pairs of 5C datasets, we tested for the difference in the ranks of interaction frequencies obtained between two pairs of genomic regions. To this end, we first bin the 5C data into 30Kb by 30Kb bins that overlap each other by 10Kb. For each of these

bins, we pool all 5C interactions that fall within it across the three biological replicates for that strain. We use this list of pooled values for the first strain and perform a two-tailed rank sum test with the list of corresponding values for the second strain. We correct for multiple testing and plot the $\log_2$ of the median of strain 1 divided by the median of strain 2 for that bin in the heat map if the corrected p-value from the test between these two bins is less than or equal to a 5% FDR.

Insertion of *Lac*O and *Tet*O arrays at RE and CENIII:

Plasmids used to integrate the *Tet and Lac* operators and express the respective TetR-mRFP and GFP-lacI repressor fusion proteins are as described (Lassadi and Bystricky 2011). All integrations were verified by colony PCR, and yielded a single fluorescent spot in all cells imaged. PCR-amplified genomic fragments with the indicated SGD coordinates were used to target specific genomic loci for inserting the operator repeats: ChrIII:26,969-27,433 (Recombination enhancer), ChrIII:115,899-116,396 (Cen).

Cell culture and Microscopy:

For live and intact cell imaging, cultures carrying two color tagged integration sites were grown exponentially in YPD to $OD_{600\ nm} <= 0.4$ (~1×10^7 cells/ml), and

rinsed in complete synthetic medium before imaging. Microscopy was performed at 25°C. Cells were spread on concave microscopy slides filled with synthetic complete 3% agarose patches for acquisition, then sealed using the VaLaP (1/3 vaseline, 1/3 lanoline, 1/3 paraffin) medium. Live microscopy was performed using an Olympus IX-81 wide-field fluorescence microscope, equipped with a CoolSNAPHQ camera (Roper Scientific) and a Polychrome V (Till Photonics), electric piezo with accurancy of 10 nm and imaged through an Olympus oil immersion objective 100X PLANAPO NA1.4. The exposure times were 300ms for both CFP and mRFP. Different stacks of 21 images each, with a step size of 0.2 µm, were taken. For 3D distance determination on G1-phase nuclei, CFP and mRFP images were automatically analyzed using SpotDistance implemented as a plug-in for ImageJ, freely available at: http://bigwww.epfl.ch/spotdistance.

CHAPTER V

Discussion and Future Directions

Possible Mechanisms of RE Function

The data presented in this thesis, for the first time, test the hypothesis that the RE modulates the conformation of Chr. III in a cell type-specific manner. The data reveals that the RE is a composite element that has multiple functions (**Fig. 5.1**). The data also show that, structurally, the RE does not behave as predicted by donor preference experiments. This does not however, contradict the donor preference data but instead provides a new perspective of that data. There is much still to be learned about the mechanism of the RE and studying this process will add to the knowledge of how cells in general control chromatin conformation.

Given the conformational data presented in this thesis and all the other molecular data presented in the literature, the mechanism of the RE is starting to take shape. In *MAT***a** cells, during G1, Mcm1p and SBF bind to the RE and initiate transcription. This process de-phases the nucleosomes in the RE, which in turn increases the accessibility of multiple Fkh1 binding sites. The increased accessibility of the RE also allows yKu70/80 to bind to the RE. The conformation

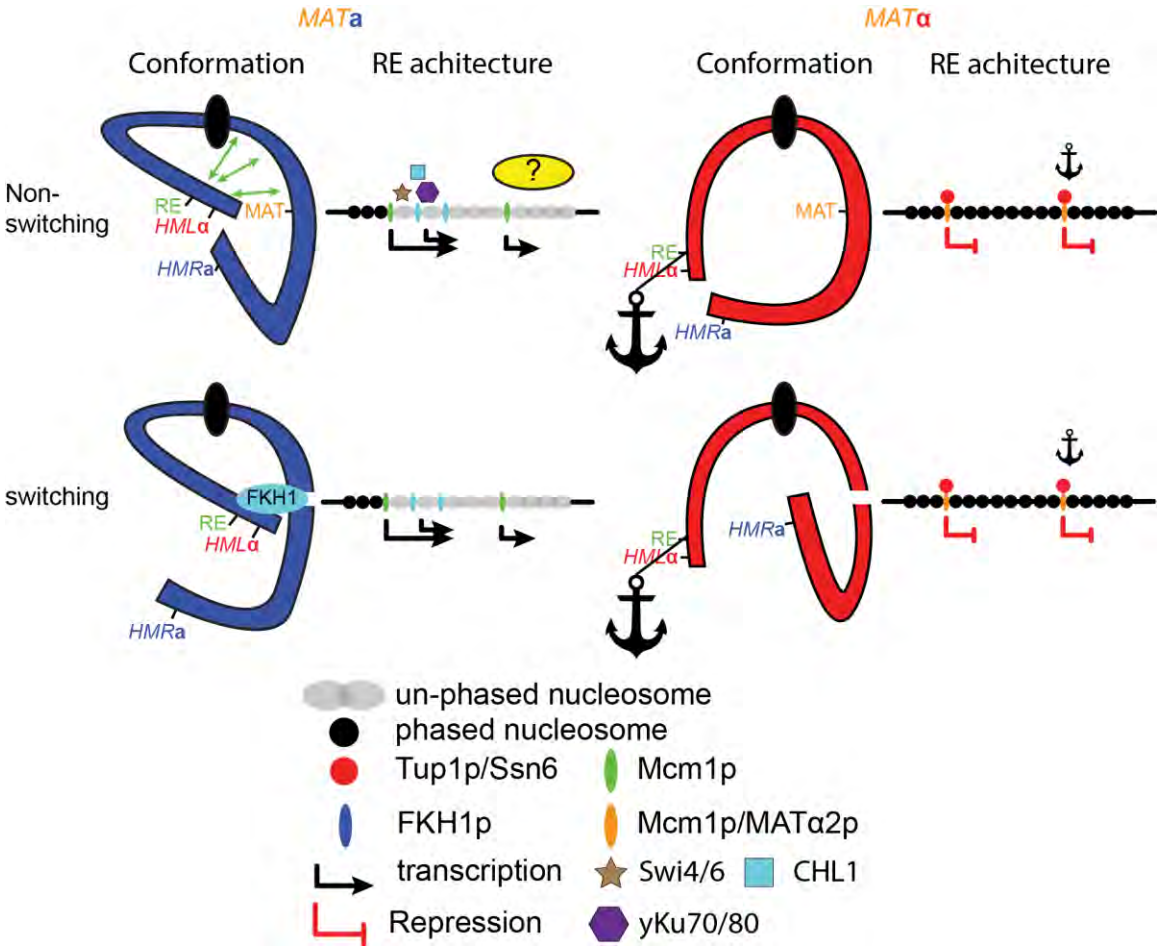


Figure 5.1. Potential Mechanism for RE's function. Schematic of the chromosome conformation and the RE architecture in both switching and non-switching conditions in both *MAT***a** and *MAT*α cells.

of the left arm of Chr. III is such that it interacts frequently with a large region of the right arm which extends from the centromere proximal region to the *MAT* locus. This process may be mediated by the action of yKu70/80 binding to *HML* and antagonizing the association of the left telomere with the nuclear periphery.

In *MAT*α cells, in G1, Mcm1p/MATα2p and Tup1/Ssn6 bind to the RE phases nucleosomes which in turn prevents Fkh1 from binding. The left arm of Chr. III is extended away from the centromere and the *MAT* locus. This may be due to the lack of yKu70/80 antagonizing the association of the left arm with the nuclear periphery.

When switching occurs, in *MAT***a** cells, yKu80, through an unknown mechanism, is transferred from the RE to *HML* as a signal to release the left arm from the periphery. Fkh1 bound to the RE catches the break at *MAT* through the action of its FHA domain (Li et al. 2012). In *MAT*α cells, the yKu70/80 signal is never sent from the RE to *HML* so the left arm may remain associated with the periphery. However, yKu70/80 at *HMR* antagonizes its association with the nuclear periphery. Since *HMR* and *HML* are close to one another on the same chromosome arm, they interact with each other relatively frequently, so a catching mechanism is not necessary.

Future Directions

There are several pieces of evidence presented by the data in this thesis that are missing from this model. First, it is unclear what mechanism would allow the RE to mediate *HML-HMR* interactions. The 5C data presented in chapter IV as well as microscopy data from others, show that *HML* and *HMR* interact much more frequency than other sub-telomeric regions on short arm chromosomes do. The 5C data also show that deleting the right portion of the RE decreases the frequency of *HML-HMR* interaction in both mating types. It appears that the association of the *HM* loci with the periphery is an indirect effect due to the affinity of telomeres to associate to the nuclear periphery (Bystricky et al. 2009b). Also, *HML* and *HMR* have been observed to interact with each other independently of their association with the periphery (Miele et al. 2009). Therefore, association of *HML* and *HMR* is not a passive phenomenon due to independent anchoring of the RE to the nuclear periphery.

yKu80 potentially acts as a messenger from the RE to *HML* during switching in *MAT***a** cells. This is likely not the mechanism that allows the RE to direct these interactions because *HML-HMR* interactions are observed independent of switching.

It could be that the RE physically interacts with the HM loci and helps hold them together. This may not have been observed before because the RE is so close to *HML* that a strong interaction between *HML* and *HMR* might occlude the observation of a strong interaction with the RE. This hypothesis could be tested by moving the RE to another chromosome and measuring the interactions

between *HML*, *HMR*, and RE using either live cell imaging or 3C based technologies.

It is puzzling that both the deletion of the left portion of the RE and deletion of Fkh1 had no effect on the conformation in *MAT***a** cells even though these parts are responsible for the majority of left arm usage during switching in *MAT***a** cells. This implies that the mechanism that establishes the conformation of Chr. III in *MAT***a** cells is distinct from the mechanism that directs left arm usage during switching. This is in line with the hypothesis that the left portion of the RE captures the break at *MAT* through the action of Fkh1 during switching in *MAT***a** cells. Since the cells in the 5C experiments are not switching, no conformational phenotype was observed. However if one were to measure conformation, as a time course, in a population of *MAT***a** cells that were synchronized in G1 by alpha-factor arrest and then induced for HO expression, there would likely be a change in conformation in these mutants. In the wild type case, if this mechanism is correct, one would likely observe a sharp peak of interaction form between RE and *MAT* until about 30-45 mins after HO induction and then it would decrease again. At some point after *MAT***α** has started to be expressed and *MAT***a**1p has sufficiently degraded, the conformation should then become the wild type *MAT***α** conformation. Deletion of RE-left or Fkh1 should result in a less intense interactions between *HMR* and *MAT* and may persist for longer than the *HML-MAT* peak in the wild type time course. This is because this process is less efficient than switching with *HML*. Since switching in these two strains

would keep the cell in the *MAT***a** state, then the *MAT***a** conformation of chr. III should persist after switching.

It is surprising that the *re⊗–left* and *fkh1⊗* mutants have a significant change of conformation in *MAT***α** cells in which the RE-left is "inactive". Given the model that the RE is in a repressive state in *MAT***α** cells, then loss of the Mcm1p/*MAT*α2p repressor binding site at DPS1 in the *re⊗–left* mutant might partially activate the RE. This is in line with the hypothesis that repression of the RE is a cooperative effect between DPS1 and DPS2 and that although DPS1 is stronger than DPS2, they are both required for full repression. However, the fact that the *fkh1⊗* mutant has a very similar conformation as the *re⊗–left* mutant suggests that the effect of this mutant is due to deleting the Fkh1 binding sites. The next experiment to do would be to delete combinations of the Fkh1 binding sites in the RE and look for a change in conformation. Since Fkh1 is a transcription factor, it may regulate the expression of a component that is involved in the conformation of the left arm of Chr. III, which may bind to the RE-left directly or indirectly and therefore, deleting RE-left and Fkh1 could have similar phenotypes because of an indirect effect.

Another question is how does the right portion of the RE affect the structure of the left arm in both mating types? The right part of the RE has been neglected in the switching literature since it only modestly decreases left arm usage during switching when deleted, and thus is non-essential. Therefore,

there is not much of a foundation to build on from the literature to explore the mechanism of how the right part of the RE may confer a change in conformation of the left arm in both mating types. A series of deletions of the right part of the RE would be the first experiment to do. These deletions should be carried out in a context other than the left arm of Chr. III to ensure that there is no contribution of any other parts of the RE to the phenotype. For instance these deletions should be carried out on the left arm of Chr. VIII since it is very similar to the size of the left arm of Chr. III and therefore any effect due to difference in arm length can be ruled out. ARS304 in the right part of the RE is very conspicuous since ARS's are used in *HML* and *HMR* to help silence these regions. The deletion series should make sure to include a whole deletion of ARS304 in order to assess its role in the overall function of RE-right. ChIP-qPCR should also be used to probe for differences in the binding of other factors to RE-right between the two mating types.

The Mating type Switching System Leverages Chromosome Conformation and Nuclear Organization at Multiple Levels

The data presented in this thesis, for the first time, provides a structural context for the mating type switching system and reveals that this complex system has evolved to leverage chromosome conformation at multiple levels. The first level of conformation that this system uses is that the entire mating type

switching cassette is located on one chromosome. On average intra-chromosomal interactions are orders of magnitude higher than the average inter-chromosomal interactions. Therefore the probability of interaction between the break at *MAT* and the *HM* loci during switching is much higher than it would be if they were located on a different chromosome. Next, the cassette is located on a chromosome that has two short arms. Short chromosome arms in yeast are more bunched together than long arms due to volume exclusion effects (Tjong et al. 2012). Also, short chromosome arms have higher interaction frequencies between pairs of loci on them because the genomic distances are small. These two facts about short chromosomes further increases the probability that the *HM* loci would interact with *MAT* during switching. The positioning of *HML*, *MAT*, and *HMR* or chromosome III is such that it makes both *HML* and *HMR* equally likely to interact with *MAT* in the absence of a molecular mechanism. *HML* is much further from *MAT* than *HMR* is, but the kink at the centromere increases the probability that *HML* would interact with *MAT*. The last few mechanisms discussed are “passive”, which means that the general arrangement of yeast chromosomes and random walk movement of the chromatin polymer is sufficient to bring loci together frequently. In addition, “active” mechanisms also play a role. An active mechanism is one in which specific factors play a role in establishing conformational differences. The interactions between *HML* and *HMR* make the right arm of Chr. III fold back on itself because the left arm is much shorter than the right arm. The folding back further increases the contact

probability of *HMR* with *MAT*. This is probably why it seems that *HMR* is the default donor when the RE is deleted in *MAT***a** cells. The action of the RE, although the exact mechanism isn't clear, brings the left arm of chromosome III closer to the *MAT* locus in *MAT***a** cells. The work described in this thesis shows that the action of the RE, not known yet in molecular terms, brings the left arm of chromosome III closer to the *MAT* locus in *MAT***a** cells even before switching initiates. It seems likely that yKu70 plays a role in releasing the arm from its sequestration at the periphery where it can then explore the nuclear volume more freely. Lastly, it has been hypothesized that the action of Fkh1p at the RE, during switching, forms a bridge between the RE and the break at *MAT* in *MAT***a** cells. This will increase the affinity of the RE with the break at *MAT* which will position *HML* near the *MAT* locus in 3D space.

All of these instances illustrate that chromosome conformation and general nuclear organization is indeed leveraged by biological systems in different ways. This also illustrates the importance of studying the role of nuclear organization as it pertains to the mating type switching cassette. Through studying this system we can better understand the general principles and mechanisms that manipulated chromosome conformation.

Mating Type Switching in the Context of General Chromatin Conformation

Studying chromosome conformation in budding yeast has added to our understanding of how cells can modulate chromosome structure as part of their biology. The RE employs two independent mechanisms to manipulate chromosome conformation during mating type switching. The first mechanism is still somewhat hypothetical because it has not been shown directly. This mechanism is the RE's ability to catch sites of DNA damage through the mechanism of Fkh1 binding to the RE and the FHA domain of Fkh1 binding to sites of DNA damage. This is an effort to target the nearby donor to the broken *MAT* locus. Evidence for this mechanism lies in the fact that the RE's effectiveness diminishes the further the donor is from the RE. This mechanism illustrates that cells can increase the probability that two loci interacting with each other by the use of a protein that binds to both loci. The second mechanism is the right portion of the RE's ability to modulate chromosome conformation in the two mating types before switching initiates. The exact mechanism of this is not known but may involve tethering of the RE to the nuclear periphery in *MAT*α cells (Avşaroğlu et al. 2014). This mechanism serves to increase the probability of interaction between *HML* and *MAT* in *MAT***a** cells and to decrease the probability of these interactions in *MAT*α cells.

These phenomenon are likely to be at play in higher eukaryotes at the length scale of TADs. Within TADs, genes and regulatory elements are generally brought closer together by a mechanism that is yet to be determined. It is

conceivable that these genes and regulator elements also bind cis acting factors that increase the affinity of these specific pairs of loci for one another.

Bibliography

Agmon N, Liefshitz B, Zimmer C, Fabre E, Kupiec M. 2013. Effect of nuclear architecture on the efficiency of double-strand break repair. *Nat Cell Biol* **15**: 694-699.

Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, Trono D, Spitz F, Duboule D. 2013. A Switch Between Topological Domains Underlies HoxD Genes Collinearity in Mouse Limbs. *Science* **340**.

Avşaroğlu B, Bronk G, Gordon-Messer S, Ham J, Bressan DA, Haber JE, Kondev J. 2014. Effect of Chromosome Tethering on Nuclear Organization in Yeast. *Plos One* **9**: e102474.

Bau D, Marti-Renom MA. 2011. Structure determination of genomic domains by satisfaction of spatial restraints. *Chromosome Research* **19**: 25-35.

Bau D, Sanyal A, Lajoie BR, Capriotti E, Byron M, Lawrence JB, Dekker J, Marti-Renom MA. 2011. The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. *Nat Struct Mol Biol* **18**: 107-+.

Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. 2012. Hi–C: A comprehensive technique to capture the conformation of genomes. *Methods* **58**: 268-276.

Berg HC. 1983. *Random Walks in Biology*. Princeton University Press.

Berger AB, Cabal GG, Fabre E, Duong T, Buc H, Nehrbass U, Olivo-Marin J-C, Gadal O, Zimmer C. 2008. High-resolution statistical mapping reveals gene territories in live yeast. *Nat Meth* **5**: 1031-1037.

Bobola N, Jansen R-P, Shin TH, Nasmyth K. 1996. Asymmetric Accumulation of Ash1p in Postanaphase Nuclei Depends on a Myosin and Restricts Yeast Mating-Type Switching to Mother Cells. *Cell* **84**: 699-709.

Bressan DA, Vazquez J, Haber JE. 2004a. Mating type-dependent constraints on the mobility of the left arm of yeast chromosome III. *J Cell Biol* **164**: 361-371.

Buchman AR, Kimmerly WJ, Rine J, Kornberg RD. 1988. Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences, and telomeres in Saccharomyces cerevisiae. *Mol Cell Biol* **8**: 210-225.

Bystricky K, Van Attikum H, Montiel M-D, Dion V, Gehlen L, Gasser SM. 2009a. Regulation of Nuclear Positioning and Dynamics of the Silent Mating Type Loci by the Yeast Ku70/Ku80 Complex. *Mol Cell Biol* **29**: 835-848.

Cabal GG, Genovesio A, Rodriguez-Navarro S, Zimmer C, Gadal O, Lesne A, Buc H, Feuerbach-Fournier F, Olivo-Marin J-C, Hurt EC et al. 2006. SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature* **441**: 770-773.

Carter D, Chakalova L, Osborne CS, Dai Y-f, Fraser P. 2002. Long-range chromatin regulatory interactions in vivo. *Nat Genet* **32**: 623-626.

Coic E, Martin J, Ryu T, Tay SY, Kondev J, Haber JE. 2011. Dynamics of Homology Searching During Gene Conversion in Saccharomyces cerevisiae Revealed by Donor Competition. *Genetics* **189**: 1225-1233.

Coic E, Richard GF, Haber JE. 2006. Saccharomyces cerevisiae donor preference during mating-type switching is dependent on chromosome architecture and organization. *Genetics* **173**: 1197-1206.

Coïc E, Sun K, Wu C, Haber JE. 2006. Cell Cycle-Dependent Regulation of Saccharomyces cerevisiae Donor Preference during Mating-Type Switching by SBF (Swi4/Swi6) and Fkh1. *Mol Cell Biol* **26**: 5470-5480.

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

Dehghani H, Dellaire G, Bazett-Jones DP. 2005. Organization of chromatin in the interphase mammalian cell. *Micron* **36**: 95-108.

Dekker J. 2006. The three 'C's of chromosome conformation capture: controls, controls, controls. *Nat Methods* **3**: 17-21.

Dekker J, Marti-Renom MA, Mirny LA. 2013. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet* **14**: 390-403.

Dekker J, Rippe K, Dekker M, Kleckner N. 2002a. Capturing chromosome conformation. *Science* **295**: 1306-1311.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012a. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **advance online publication**.

Dostie J, Dekker J. 2007. Mapping networks of physical interactions between genomic elements using 5C technology. *Nature Protocols* **2**: 988-1002.

Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C et al. 2006a. Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Research* **16**: 1299-1309.

Doyle B, Fudenberg G, Imakaev M, Mirny LA. 2014. Chromatin Loops as Allosteric Modulators of Enhancer-Promoter Interactions. *Plos Comput Biol* **10**: e1003867.

Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, Lee C, Shendure J, Fields S, Blau CA, Noble WS. 2010a. A three-dimensional model of the yeast genome. *Nature* **465**: 363-367.

Ercan S, Reese JC, Workman JL, Simpson RT. 2005. Yeast recombination enhancer is stimulated by transcription activation. *Mol Cell Biol* **25**: 7976-7987.

Ercan S, Simpson RT. 2004. Global chromatin structure of 45,000 base pairs of chromosome III in a- and alpha-cell yeast and during mating-type switching. *Mol Cell Biol* **24**: 10026-10035.

Feng S, Cokus Shawn J, Schubert V, Zhai J, Pellegrini M, Jacobsen Steven E. 2014. Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis. *Mol Cell* **55**: 694-707.

Ferraiuolo MA, Sanyal A, Naumova N, Dekker J, Dostie J. 2012a. From cells to chromatin: Capturing snapshots of genome organization with 5C technology. *Methods* **58**: 255-267.

Fishmanlobell J, Haber JE. 1992. REMOVAL OF NONHOMOLOGOUS DNA ENDS IN DOUBLE-STRAND BREAK RECOMBINATION - THE ROLE OF THE YEAST ULTRAVIOLET REPAIR GENE RAD1. *Science* **258**: 480-484.

Fox CA, Ehrenhofer-Murray AE, Loo S, Rine J. 1997. The Origin Recognition Complex, SIR1, and the S Phase Requirement for Silencing. *Science* **276**: 1547-1551.

Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH et al. 2009a. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**: 58-64.

Galgoczy DJ, Cassidy-Stone A, Llinás M, O'Rourke SM, Herskowitz I, DeRisi JL, Johnson AD. 2004. Genomic dissection of the cell-type-specification circuit in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences* **101**: 18069-18074.

Gheldof N, Smith EM, Tabuchi TM, Koch CM, Dunham I, Stamatoyannopoulos JA, Dekker J. 2010. Cell-type-specific long-range looping interactions identify distant regulatory elements of the CFTR gene. *Nucleic Acids Res* **38**: 4325-4336.

Gotta M, Laroche T, Formenton A, Maillet L, Scherthan H, Gasser SM. 1996. The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type Saccharomyces cerevisiae. *The Journal of Cell Biology* **134**: 1349-1363.

Griffith J, Hochschild A, Ptashne M. 1986. DNA loops induced by cooperative binding of [lambda] repressor. *Nature* **322**: 750-752.

Grob S, Schmid M, Luedtke N, Wicker T, Grossniklaus U. 2013. Characterization of chromosomal architecture in Arabidopsis by chromosome conformation capture. *Genome Biol* **14**: R129.

Grob S, Schmid Marc W, Grossniklaus U. 2014. Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. *Mol Cell* **55**: 678-693.

Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W et al. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**: 948-951.

Haber JE. 1992. Mating-type gene switching in Saccharomyces cerevisiae. *Trends in Genetics* **8**: 446-452.

Haber JE. 1998a. A locus control region regulates yeast recombination. *Trends Genet* **14**: 317-321.
Haber JE. 1998b. MATING-TYPE GENE SWITCHING IN SACCHAROMYCES CEREVISIAE. *Annual Review of Genetics* **32**: 561-599.
Haber JE. 2012a. Mating-Type Genes and MAT Switching in Saccharomyces cerevisiae. *Genetics* **191**: 33-64.
Haber JE, George JP. 1979. A Mutation That Permits the Expression of Normally Silent Copies of Mating-Type Information in SACCHAROMYCES CEREVISIAE. *Genetics* **93**: 13-35.
Haber JE, Ray BL, Kolb JM, White CI. 1993. RAPID KINETICS OF MISMATCH REPAIR OF HETERODUPLEX DNA THAT IS FORMED DURING RECOMBINATION IN YEAST. *P Natl Acad Sci USA* **90**: 3363-3367.
Hagen DC, Bruhn L, Westby CA, Sprague GF. 1993. Transcription of alpha-specific genes in Saccharomyces cerevisiae: DNA sequence requirements for activity of the coregulator alpha 1. *Mol Cell Biol* **13**: 6866-6875.
Hicks WM, Yamaguchi M, Haber JE. 2011. Real-time analysis of double-strand DNA break repair by homologous recombination. *P Natl Acad Sci USA* **108**: 3108-3115.
Horike S, Cai S, Miyano M, Cheng JF, Kohwi-Shigematsu T. 2005. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet* **37**: 31-40.
Houston P, Simon PJ, Broach JR. 2004. The Saccharomyces cerevisiae recombination enhancer biases recombination during interchromosomal mating-type switching but not in interchromosomal homologous recombination. *Genetics* **166**: 1187-1197.
Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, Dekker J, Mirny LA. 2012. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Meth* **9**: 999-1003.
Ivanov EL, Sugawara N, FishmanLobell J, Haber JE. 1996. Genetic requirements for the single-strand annealing pathway of double-strand break repair in Saccharomyces cerevisiae. *Genetics* **142**: 693-704.
Ivanov EL, Sugawara N, White CI, Fabre F, Haber JE. 1994. MUTATIONS IN XRS2 AND RAD50 DELAY BUT DO NOT PREVENT MATING-TYPE SWITCHING IN SACCHAROMYCES-CEREVISIAE. *Mol Cell Biol* **14**: 3414-3425.
Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen C-A, Schmitt AD, Espinoza CA, Ren B. 2013. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**: 290-294.
Jin Q-w, Trelles-Sticken E, Scherthan H, Loidl J. 1998a. Yeast Nuclei Display Prominent Centromere Clustering That Is Reduced in Nondividing Cells and in Meiotic Prophase. *The Journal of Cell Biology* **141**: 21-29.
Jin QW, Fuchs J, Loidl J. 2000a. Centromere clustering is a major determinant of yeast interphase nuclear organization. *Journal of Cell Science* **113**: 1903-1912.

Kaplun L, Ivantsiv Y, Bakhrat A, Raveh D. 2003. DNA Damage Response-mediated Degradation of Ho Endonuclease via the Ubiquitin System Involves Its Nuclear Export. *Journal of Biological Chemistry* **278**: 48727-48734.

Kimmerly W, Buchman A, Kornberg R, Rine J. 1988. Roles of two DNA-binding factors in replication, segregation and transcriptional repression mediated by a yeast silencer. *The EMBO Journal* **7**: 2241-2253.

Kind J, Pagie L, Ortabozkoyun H, Boyle S, de Vries Sandra S, Janssen H, Amendola M, Nolen Leisha D, Bickmore Wendy A, van Steensel B. 2013. Single-Cell Dynamics of Genome-Nuclear Lamina Interactions. *Cell* **153**: 178-192.

Klar AJS. 1987. Determination of the yeast cell lineage. *Cell* **49**: 433-435.

Knott Simon RV, Peace Jared M, Ostrow AZ, Gan Y, Rex Alexandra E, Viggiani Christopher J, Tavaré S, Aparicio Oscar M. 2012. Forkhead Transcription Factors Establish Origin Timing and Long-Range Clustering in S. cerevisiae. *Cell* **148**: 99-111.

Kobayashi T, Heck DJ, Nomura M, Horiuchi T. 1998. Expansion and contraction of ribosomal DNA repeats in Saccharomyces cerevisiae: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Genes & Development* **12**: 3821-3830.

Lajoie BR, van Berkum NL, Sanyal A, Dekker J. 2009a. My5C: web tools for chromosome conformation capture studies. *Nat Methods* **6**: 690-691..

Lang GI, Parsons L, Gammie AE. 2013. Mutation rates, spectra, and genome-wide distribution of spontaneous mutations in mismatch repair deficient yeast. *G3 (Bethesda)* **3**: 1453-1465.

Lassadi I, Bystricky K. 2011. Tracking of single and multiple genomic loci in living yeast cells. *Methods Mol Biol* **745**.

Le Dily F, Baù D, Pohl A, Vicent GP, Serra F, Soronellas D, Castellano G, Wright RHG, Ballare C, Filion G et al. 2014. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes & Development* **28**: 2151-2162.

Le TBK, Imakaev MV, Mirny LA, Laub MT. 2013. High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. *Science* **342**: 731-734.

Li J, Coïc E, Lee K, Lee C-S, Kim J-A, Wu Q, Haber JE. 2012. Regulation of Budding Yeast Mating-Type Switching Donor Preference by the FHA Domain of Fkh1. *Plos Genet* **8**: e1002630.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO et al. 2009a. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**: 289-293.

McCord RP, Nazario-Toole A, Zhang H, Chines PS, Zhan Y, Erdos MR, Collins FS, Dekker J, Cao K. 2013. Correlated alterations in genome organization, histone methylation, and DNA–lamin A/C interactions in Hutchinson-Gilford progeria syndrome. *Genome Research* **23**: 260-269.

Meister P, Gehlen LR, Varela E, Kalck V, Gasser SM. 2010. Chapter 21 - Visualizing Yeast Chromosomes and Nuclear Architecture. in *Methods in Enzymology* (eds. W Jonathan, G Christine, RF Gerald), pp. 535-567. Academic Press.

Miele A, Bystricky K, Dekker J. 2009. Yeast Silent Mating Type Loci Form Heterochromatic Clusters through Silencer Protein-Dependent Long-Range Interactions. *Plos Genet* **5**.

Miller AM, Nasmyth KA. 1984. Role of DNA replication in the repression of silent mating type loci in yeast. *Nature* **312**: 247-251.

Mirny LA. 2011. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Research* **19**: 37-51.

Miyazaki T, Bressan DA, Shinohara M, Haber JE, Shinohara A. 2004. In vivo assembly and disassembly of Rad51 and Rad52 complexes during double-strand break repair. *Embo Journal* **23**: 939-949.

Mizuguchi T, Fudenberg G, Mehta S, Belton J-M, Taneja N, Folco HD, FitzGerald P, Dekker J, Mirny L, Barrowman J et al. 2014. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. *Nature* **advance online publication**.

Monneron A, Bernhard W. 1969. Fine structural organization of the interphase nucleus in some mammalian cells. *Journal of Ultrastructure Research* **27**: 266-288.

Nasmyth K. 1987. The determination of mother cell-specific mating type of switching in yeast by a specific regulator of HO transcription. *The EMBO Journal* **6**: 243-248.

Nasmyth K. 1993. Regulating the HO endonuclease in yeast. *Current Opinion in Genetics & Development* **3**: 286-294.

Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny LA, Dekker J. 2013. Organization of the Mitotic Chromosome. *Science* **342**: 948-953.

Naumova N, Smith EM, Zhan Y, Dekker J. 2012. Analysis of long-range chromatin interactions using Chromosome Conformation Capture. *Methods* **58**: 192-203.

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J et al. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **advance online publication**.

Ostrow AZ, Nellimoottil T, Knott SRV, Fox CA, Tavaré S, Aparicio OM. 2014. Fkh1 and Fkh2 Bind Multiple Chromosomal Elements in the <italic>S. cerevisiae</italic> Genome with Distinct Specificities and Cell Cycle Dynamics. *Plos One* **9**: e87647.

Oza P, Jaspersen SL, Miele A, Dekker J, Peterson CL. 2009. Mechanisms that regulate localization of a DNA double-strand break to the nuclear periphery. *Genes & Development* **23**: 912-927.

Palstra R-J, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W. 2003. The [beta]-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* **35**: 190-194.

Patterson EE, Fox CA. 2008. The Ku Complex in Silencing the Cryptic Mating-Type Loci of Saccharomyces cerevisiae. *Genetics* **180**: 771-783.

Ravindra A, Weiss K, Simpson RT. 1999. High-Resolution Structural Analysis of Chromatin at Specific Loci: Saccharomyces cerevisiae Silent Mating-Type Locus HMR a. *Mol Cell Biol* **19**: 7944-7950.

Ray BL, White CI, Haber JE. 1991. HETERODUPLEX FORMATION AND MISMATCH REPAIR OF THE STUCK MUTATION DURING MATING-TYPE SWITCHING IN SACCHAROMYCES-CEREVISIAE. *Mol Cell Biol* **11**: 5372-5380.

Rine J, Herskowitz I. 1987. Four Genes Responsible for a Position Effect on Expression from HML and HMR in Saccharomyces cerevisiae. *Genetics* **116**: 9-22.

Ruan C, Workman JL, Simpson RT. 2005. The DNA repair protein yKu80 regulates the function of recombination enhancer during yeast mating type switching. *Mol Cell Biol* **25**: 8476-8485.

Rustchenko EP, Curran TM, Sherman F. 1993. Variations in the number of ribosomal DNA units in morphological mutants and normal strains of Candida albicans and in normal strains of Saccharomyces cerevisiae. *J Bacteriol* **175**: 189-199.

Sanyal A, Lajoie BR, Jain G, Dekker J. 2012. The long-range interaction landscape of gene promoters. *Nature* **489**: 109-113.

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. 2012. Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell* **148**: 458-472.

Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W. 2006a. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**: 1348-1354.

Solovei I, Cavallo A, Schermelleh L, Jaunin F, Scasselati C, Cmarko D, Cremer C, Fakan S, Cremer T. 2002. Spatial preservation of nuclear chromatin architecture during three-dimensional fluorescence in situ hybridization (3D-FISH). *Exp Cell Res* **276**: 10-23.

Spilianakis CG, Flavell RA. 2004. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol* **5**: 1017-1027.

Strathern JN, Herskowitz I. 1979. Asymmetry and directionality in production of new cell types during clonal growth: the switching pattern of homothallic yeast. *Cell* **17**: 371-381.

Sugawara N, Haber JE. 2006. Repair of DNA double strand breaks: In vivo biochemistry. in *DNA Repair, Pt A* (eds. J Campbell, P Modrich), pp. 416-429.

Sugawara N, Ivanov EL, Fishmanlobell J, Ray BL, Wu X, Haber JE. 1995. DNA STRUCTURE-DEPENDENT REQUIREMENTS FOR YEAST RAD GENES IN GENE CONVERSION. *Nature* **373**: 84-86.

Sun K, Coïc E, Zhou Z, Durrens P, Haber JE. 2002a. Saccharomyces forkhead protein Fkh1 regulates donor preference during mating-type switching through the recombination enhancer. *Genes Dev* **16**: 2085-2096.

Symmons O, Uslu VV, Tsujimura T, Ruf S, Nassari S, Schwarzer W, Ettwiller L, Spitz F. 2014. Functional and topological characteristics of mammalian regulatory domains. *Genome Research* **24**: 390-400.

Szeto L, Broach JR. 1997. Role of alpha2 protein in donor locus selection during mating type interconversion. *Mol Cell Biol* **17**: 751-759.

Szeto L, Fafalios MK, Zhong H, Vershon AK, Broach JR. 1997a. Alpha2p controls donor preference during mating type interconversion in yeast by inactivating a recombinational enhancer of chromosome III. *Genes Dev* **11**: 1899-1911.

Tatchell K, Nasmyth KA, Hall BD, Astell C, Smith M. 1981. In vitro mutation analysis of the mating-type locus in yeast. *Cell* **27**: 25-35.

Therizols P, Duong T, Dujon B, Zimmer C, Fabre E. 2010. Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proceedings of the National Academy of Sciences* **107**: 2025-2030.

Tjong H, Gong K, Chen L, Alber F. 2012. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Research*.

Tolhuis B, Palstra R-J, Splinter E, Grosveld F, de Laat W. 2002. Looping and Interaction between Hypersensitive Sites in the Active β-globin Locus. *Mol Cell* **10**: 1453-1465.

Trelles-Sticken E, Dresser ME, Scherthan H. 2000. Meiotic Telomere Protein Ndj1p Is Required for Meiosis-Specific Telomere Distribution, Bouquet Formation and Efficient Homologue Pairing. *The Journal of Cell Biology* **151**: 95-106.

Umbarger MA, Toro E, Wright MA, Porreca GJ, Bau D, Hong SH, Fero MJ, Zhu LJ, Marti-Renom MA, McAdams HH et al. 2011. The Three-Dimensional Architecture of a Bacterial Genome and Its Alteration by Genetic Perturbation. *Mol Cell* **44**: 252-264.

Umezu K, Sugawara N, Chen C, Haber JE, Kolodner RD. 1998. Genetic analysis of yeast RPA1 reveals its multiple functions in DNA metabolism. *Genetics* **148**: 989-1005.

Vernimmen D, Gobbi MD, Sloane-Stanley JA, Wood WG, Higgs DR. 2007. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *The EMBO Journal* **26**: 2041-2051.

Weiler KS, Broach JR. 1992a. Donor locus selection during Saccharomyces cerevisiae mating type interconversion responds to distant regulatory signals. *Genetics* **132**: 929-942.

Weiler KS, Szeto L, Broach JR. 1995. Mutations affecting donor preference during mating type interconversion in Saccharomyces cerevisiae. *Genetics* **139**: 1495-1510.

Weiss K, Simpson RT. 1997. *Cell type-specific chromatin organization of the region that governs directionality of yeast mating type switching*.

Weiss K, Simpson RT. 1998. High-Resolution Structural Analysis of Chromatin at Specific Loci: Saccharomyces cerevisiae Silent Mating Type Locus HMLα. *Mol Cell Biol* **18**: 5392-5403.

White CI, Haber JE. 1990. INTERMEDIATES OF RECOMBINATION DURING MATING TYPE SWITCHING IN SACCHAROMYCES-CEREVISIAE. *Embo Journal* **9**: 663-673.

Williamson I, Eskeland R, Lettice LA, Hill AE, Boyle S, Grimes GR, Hill RE, Bickmore WA. 2012. Anterior-posterior differences in HoxD chromatin topology in limb development. *Development* **139**: 3157-3167.

Wu C, Weiss K, Yang C, Harris MA, Tye B-K, Newlon CS, Simpson RT, Haber JE. 1998a. Mcm1 regulates donor preference controlled by the recombination enhancer in Saccharomyces mating-type switching. *Genes & Development* **12**: 1726-1737.

Wu X, Haber JE. 1995a. MATa donor preference in yeast mating-type switching: activation of a large chromosomal region for recombination. *Genes & Development* **9**: 1922-1932.

Wu XH, Haber JE. 1996b. A 700 bp cis-acting region controls mating-type dependent recombination along the entire left arm of yeast chromosome III. *Cell* **87**: 277-285.

Wu XH, Moore JK, Haber JE. 1996. Mechanism of MAT alpha donor preference during mating-type switching of Saccharomyces cerevisiae. *Mol Cell Biol* **16**: 657-668.

Yaffe E, Tanay A. 2011b. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat Genet* **43**: 1059-1065.

Zhang Y, McCord Rachel P, Ho Y-J, Lajoie Bryan R, Hildebrand Dominic G, Simon Aline C, Becker Michael S, Alt Frederick W, Dekker J. 2012a. Spatial Organization of the Mouse Genome and Its Role in Recurrent Chromosomal Translocations. *Cell* **148**: 908-921.

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

Zhong H, McCord R, Vershon AK. 1999. Identification of Target Sites of the α2–Mcm1 Repressor Complex in the Yeast Genome. *Genome Research* **9**: 1040-1047.

Zimmer C, Fabre E. 2011. Principles of chromosomal organization: lessons from yeast. *The Journal of Cell Biology* **192**: 723-733.

Zullo Joseph M, Demarco Ignacio A, Piqué-Regi R, Gaffney Daniel J, Epstein Charles B, Spooner Chauncey J, Luperchio Teresa R, Bernstein Bradley E, Pritchard Jonathan K, Reddy Karen L et al. 2012. DNA Sequence-Dependent Compartmentalization and Silencing of Chromatin at the Nuclear Lamina. *Cell* **149**: 1474-1487.