Topological and Chromatin Alterations Influencing Genome Integrity

Thesis

Mohamood Adhil Mohammed Iqbal

Student ID: G2096245

Topological and Chromatin Alterations Influencing Genome Integrity

Supervisor: Prof. Dr. Marco Foiani

External Supervisor: Prof. Dr. Antony M Carr

Open University PhD Programme

Fundamentals of Cancer Biology

IFOM the FIRC Institute of Molecular Oncology

October 2021

1. Abstract

DNA topoisomerases Top1 and Top2 have redundant functions in resolving topological alterations arising during replication and transcription processes. Topoisomerases assist replication forks encountering transcription units, preventing chromosome fragility by minimizing the aberrant topological events. We investigated the role of topoisomerases in supercoil accumulation across the yeast genome using biotin tagged psoralen immunoprecipitation. We found that DNA is under-wound at gene boundaries and over-wound at transcribed regions. Top1 is associated with positively supercoiled chromatin as it accompanies RNA Polymerase II (Pol2) and its chromatin association is influenced by transcription levels of the individual genes. Top2 is associated with stable negative supercoiled chromatin at the gene boundaries, and its association is not dependent on transcription. Top2 promotes transcription efficiencies by forming gene loop structure and restricts Top1 and Pol2 leakage at gene boundaries. Ablation of Top2 protein decreases the negative supercoil accumulation at gene boundaries. Expression of *E.coli* TopA in topoisomerases double mutant in yeast (*top2-1top1Δ*) significantly resolves only the negative supercoil of gene boundaries and increases the accumulation of positive supercoil. The supercoil state at gene boundaries and ORFs are crucial for nucleosome occupancy. Using Hi-C techniques, we show that, centromeres are prominently interacting with other centromeres and the inter-chromosomal centromere interactions are depleted along with cohesin protein in *top2-1top1Δ* mutant expressing *E.coli* TopA. This work therefore summarizes the role of supercoil structures in preserving higher order architecture including nucleosome formation and chromosome organization.

Thesis Dedication & Acknowledgments:

I would like to thank my parents and family members for their unconditional support and love during my PhD research. I dedicate this thesis to my wife (Nazeeha Ayaz) for being a superwoman by supporting me, simultaneously performing her PhD research and raising our son. I thank my lovely son (Mohammed Arshad) for teaching me the difficulties of raising a kid, which helped me face other difficulties with ease. Last four years being a roller coaster ride during my PhD, there were periods with good results and some with a lot of troubleshooting to make the experiments work. I am very blessed to have this experience which taught me many important life lessons like being patience, critical thinking and decision making.

I am very grateful and thankful to my supervisor Prof. Dr. Marco Foiani for providing me this wonderful opportunity to work on a fascinating topic, providing the cutting-edge facilities to perform the experiment and importantly guiding me in the right scientific direction. He made me welcome and comfortable right from the first interaction. I thank my external supervisor Prof Dr. Tony Carr from University of Sussex for his guidance during my PhD research.

A special thanks to Dr. Yathish Jagadheesh Achar for his teaching, endless support and co-supervising my PhD research. Our first interaction dates back to 2014 where he motivated me to pursue research and since then we had countless discussions which helped me in my personal life and professional career.

I am very thankful to all Marco Foiani group members for helping me and providing an amazing scientific and social environment. I thank Dr. Chiara Lucca (Mother of Yeast) for sharing her knowledge on budding yeast, helping me in many personal things and sitting next to me and tolerating me for the last four years. I thank Dr. Ramveer Choudhary for helping in experiments and also sharing his knowledge. I thank all the facilities especially IFOM-Cogentech sequencing facility for their wonderful support. I thank IFOM welcome office members Mio Sumie and Marina Properzi for their support in all the documentation work which makes life easier for people coming from other countries. A big thanks to the whole IFOM institute for providing the international work environment, where I thoroughly enjoyed the last four years during my PhD research.

Finally, I would like to thank yeast organism for tirelessly working for my project by consuming few liters of media.

Table of Contents

Figure Index

Introduction

Results

Publication by the Candidate:

Achar, Y. J., M. Adhil, R. Choudhary, N. Gilbert and M. Foiani (2020). "Negative supercoil at gene boundaries modulates gene topology." Nature **577**(7792): 701-705.

(Note: This thesis contains many figures and contents from the above published paper)

Abbreviations

3C Chromosome conformation capture

ARS Autonomously Replicating Sequence

BSA Bovine Serum Albumin

bTMP biotinylated 4,5,8-trimethylpsoralean

ChIA-PET Chromatin interaction analysis by paired-end tag sequencing

ChIP-on-CHIP Chromatin Immunoprecipitation on CHIP (Affymetrix microarray)

Chr Chromosome

ddH2O double distilled H2O

DMSO Dimethyl Sulfoxide

DSB Double Strand Break

dsDNA double stranded DNA

E.coli *Escherichia coli*

EDTA Ethylen Diammino Tetraacetic Acid

Gal Galactose

Glu Glucose

Hi-C Chromosome conformation capture technique using high through put sequencing

Micro-C Micrococcal Nuclease chromatin fragmentation Hi-C

NGS next generation sequencing

ORF Open Reading Frame

PEG Polyethylene glycol

RCF Relative Centrifugal Force

RFB Replication Fork Barrier

RNAPII RNA polymerase II

RPM Revolutions Per Minute

RT Room Temperature

SDS Sodium Dodecyl Sulphate

SGA Synthetic Genetic Array

SMC Structural Maintenance of Chromosome

ssDNA single stranded DNA

TADs Topological Associated Domains

TSS Transcription Start Site

TTS Transcription Termination Site

Wt wildtype

YNB Yeast Nitrogen base

2. Introduction

Living cells are biological factories where many molecular components are made and assembled into smaller compartments known as organelles. These organelles undergo cascades of biochemical reactions and signals that are responsible for various cellular processes like cell division and apoptosis. The nucleus is a crucial cellular component which harbors our genetic blueprint, responsible for biological processes like DNA repair, RNA synthesis, protein expression, chromatin formation, genome duplication, segregation of chromosomes etc. DNA supercoiling, a fundamental property of the DNA double helix structure that is generated during replication and transcription, plays a major role in gene expression, chromatin packaging and genome organization.

In this dissertation, the role of DNA supercoiling in transcription and higher order genome architecture using *Saccharomyces cerevisiae* as a model organism was investigated. The following section will provide an overview of the topological events arising during replication and transcription, the role of topoisomerases I and II during genome transactions and the chromatin architectural proteins that are responsible for maintaining the genome integrity.

2.1 DNA Supercoil in Eukaryotes

DNA has multiple structural forms such as B-DNA, A-DNA and Z-DNA with different functional attributes. These structures are categorized based on the number of bases and helical twist between two strands. The B-DNA is the most common form found in living cells, and in its relaxed state contains approximately 10.4 nucleotides base-pair per turn. Eukaryotic cells contain three distinct RNA polymerases (Pol I, Pol II and Pol III) that transcribe different classes of genes. The protein coding genes are transcribed by RNA polymerase II whereas the ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) are

transcribed by RNA polymerase I and III. As a consequence of transcription, the flanking region experiences helical tension which results in over-wound (Positive Supercoils) and under-wound (Negative Supercoils) DNA in front of and behind the polymerase respectively, which is known as twin domain model (Liu and Wang 1987). The topoisomerases are important components for relieving topological tensions during RNA polymerase progression by inducing single (type I) and double strand breaks (type II) respectively (Wang 2002). Some DNA regions experience more helical tension and they result in non-B-DNA structure or DNA cruciform (Murchie and Lilley 1987). However, the net state of the genome is torsionally relaxed (Sinden, Carlson et al. 1980).

The hyper accumulation of negative supercoils in non-B-DNA structures help to regulate major biological processes during genome transactions. The negative supercoils also play a vital role in keeping the promoter in open state for binding of transcription preinitiation complex (PIC), transcription factors and regulatory complex (Ljungman and Hanawalt 1995). The initiation of RNA Pol2 is facilitated by the open chromatin, where the DNA is nucleosome free (Revyakin, Liu et al. 2006).

While the twin domain model explains the supercoiling during transcription, certain genomic loci are supercoiled even before transcription activation due to certain proteins involved in organization of supercoil domains (Stewart, Herrera et al. 1990) (Gilmour and Lis 1986). In *E.coli*, it has been shown that transcription of a gene influences the transcription of other genes by inducing negative supercoil in the divergent promoter (Rhee, Opel et al. 1999). Transcription of a short upstream RNA may involve formation of DNA supercoiling (Seila, Calabrese et al. 2008). Use of a compound known as bTMP (biotinylated 4,5,8-trimethylpsoralean (TMP) that intercalates between the bases of nucleotides and is also biotin-tagged to facilitate purification (Figure 1A) demonstrates that, the transcription and topoisomerase activity alters the DNA supercoiling around transcription start sites (TSS) in mammalian cells. This work demonstrates that genomic loci bTMP can be categorized as

underwound, overwound or stable regions (Naughton, Avlonitis et al. 2013) (Kouzine, Gupta et al. 2013).

In eukaryotes, DNA topoisomerase I is known to be associated with transcribing regions to resolve DNA supercoiling ahead of the RNA polymerase. In top1 mutants, the role of Top1 is fulfilled by DNA topoisomerase II as both are involved in relaxing the positively and negatively supercoiled DNA (Liu and Wang 1987) (Gartenberg and Wang 1992). Expresion of *Escherichia coli* topoisomerase I (*E.coli* TopA) in the absence of topoisomerases I and II leads to the accumulation of positive supercoiling of intracellular DNA (Gartenberg and Wang 1992). According to the twin-domain model of transcriptional supercoiling, a moving RNA polymerase generates positive supercoils ahead and negative supercoils behind. However, the specific bacterial topoisomerase (*E.coli* TopA) relaxes only the negative supercoils leading to net accumulation of positive supercoils. This accumulation in template DNA greatly reduces mRNA synthesis (Gartenberg and Wang 1992).

Till now there are no comprehensive studies performed on a genomic scale with respect to DNA supercoiling and its role in higher order chromatin organization in the nucleus. To study the supercoil state on a genomic scale and its impact on chromatin organization, we used biotinylated 4,5,8-trimethylpsoralean (bTMP) (Naughton, Avlonitis et al. 2013), which permeates the cell and preferentially intercalates into underwound regions (negatively supercoiled regions) (Figure 1B). Psoralens intercalate in between the bases of DNA and upon irradiation with ultraviolet (UV) light (365 nm) form covalent crosslinks between pyrimidines of opposite strands. Trimethylpsoralen (TMP) is the most commonly used psoralen for chromatin studies. The biotin attached via a linker to the psoralen is used to immunoprecipitate and identify the underwound regions.

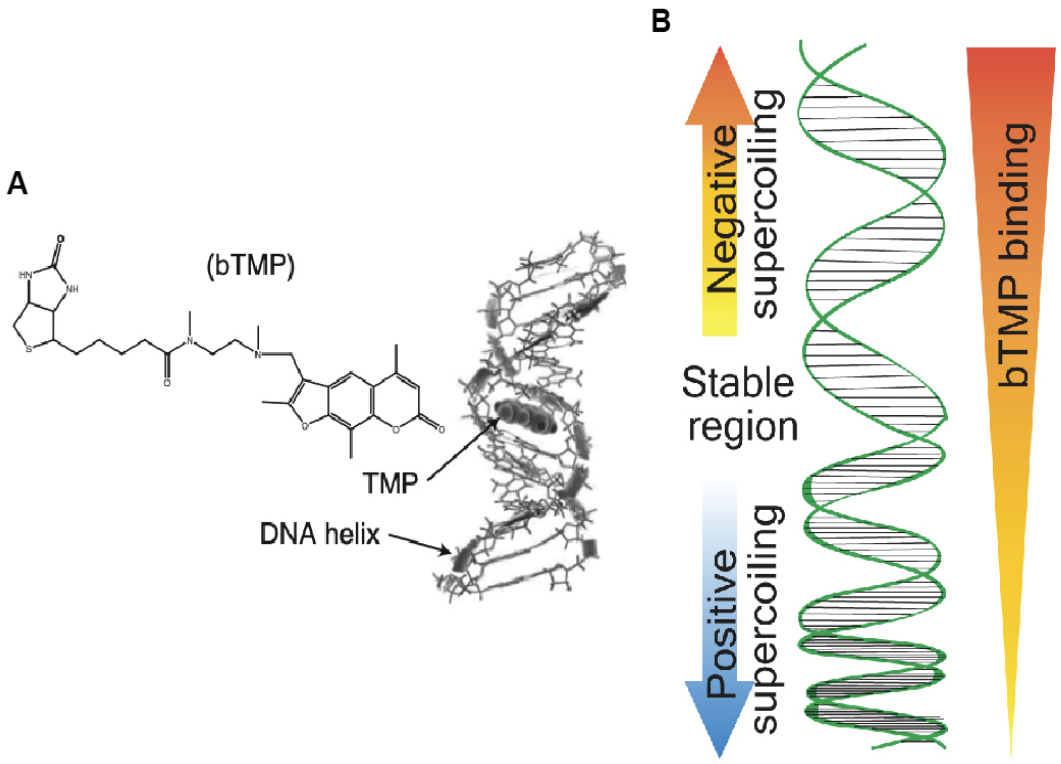


Figure 1: bTMP intercalation into double stranded DNA A) Illustration of biotin tagged psoralen (bTMP) interclates between double strand DNA (Figure adapterd from Naughton, Avlonitis et al. 2013) B) Cartoon representation of preferential binding of bTMP in underwound or negative supercoiled region.

2.2 Mechanism of type I and type II topoisomerases

DNA topoisomerases resolves the topological constraints by introducing transient breaks in DNA using a transesterification mechanism which reduces the risk of permanent DNA damage (Wang 2002). The DNA cleavage by topoisomerases is followed by formation of a covalent phosphodiester bond between a specific tyrosine residue in the catalytic site of each topoisomerase of protein and broken strand of DNA, then strand passage of either ssDNA or dsDNA, and resealing of the breaks is carried out using nucleophilic substitution. The topoisomerases can be classified into type I (TopI) and type II (TopII), where type I cleave only one strand and type II cleaves both strands to resolve topological constraints (Liu and Wang 1987). The topoisomerases do not require any specific DNA sequences (Spitzner and Muller 1988) to bind, unlike many other sequence specific chromatin binding proteins. Their activities are highly regulated both in nucleosome and nucleosome free

regions (Capranico, Jaxel et al. 1990). In nucleosomal regions, Top2 is more efficient than Top1 in relaxing nucleosomal DNA. The DNA cross-inversion mechanism of topoisomerase II is facilitated in the chromatin, which favors closeness of DNA segments, whereas the DNA strand-rotation mechanism of topoisomerase I does not efficiently relax the chromatin, thereby imposing barriers for DNA twist diffusion (Salceda, Fernandez et al. 2006). The linker DNA segments (30-90bp) between nucleosomes don't have high accumulation of supercoil. In nucleosome free regions, non-B DNA structures form due to the accumulation of negative supercoil and these structures are required for transcription initiation and elongation (Marchand, Pourquier et al. 2002).

2.3 Role of topoisomerases in transcription

Topoisomerases resolve the positive and negative DNA supercoiling that accumulates in front of or behind the transcription machinery (Liu and Wang 1987). Top1 or Top2 resolve topological constraints during transcription thereby enhancing the recruitment of RNA Pol II to promoters (Sperling, Jeong et al. 2011) (Figure 2). The topoisomerase activities are not restricted to just relaxing the DNA but are also responsible for gene expression.

Studies in yeast show that topoisomerases are required for initiating transcription of certain group of genes that are regulated by galactose and inorganic phosphate (Pedersen, Fredsoe et al. 2012) (Roedgaard, Fredsoe et al. 2015) such as GAL1, GAL2, GAL7 and GAL10. This study showed that, the topoisomerases are required for transcription initiation and not for elongation or re-initiation. GAL gene activation requires topoisomerases either directly for TATA Box Protein (TBP) binding to the TATA box, or in a step between nucleosome eviction and TBP binding.

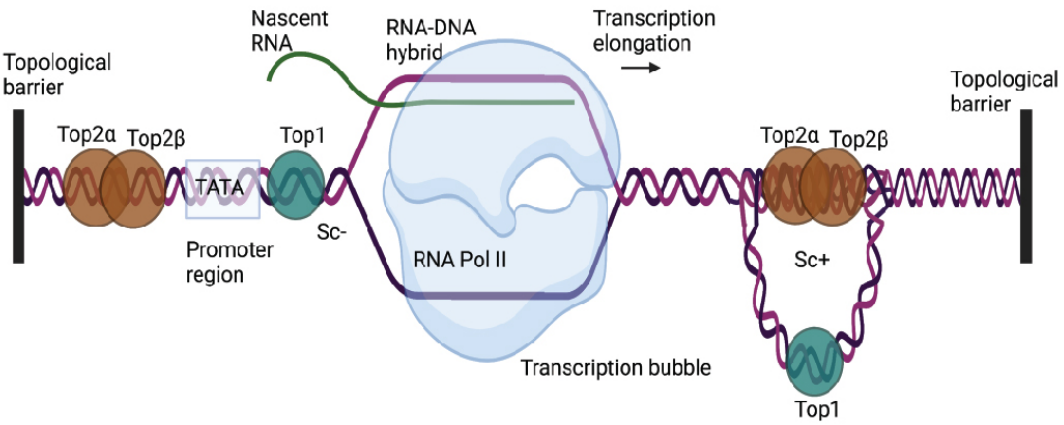


Figure 2: Topoisomerase Top1 and Top2 role in transcription. RNA Polymerase II elongation during transcription incurs topological constraints. Negative supercoil behind (Sc^-) and Positive supercoiling (Sc+) ahead of the transcription bubble which in turn obstructs further Pol II movement. Topoisomerases are involved in the relaxation of the supercoil generated during transcription. TATA-box-binding protein (TBP) at promoter are regulted by TOP1 protein and TOP2 β-mediated transient DNA double-stranded breaks are crucial transcriptional regulation.

The highly expressed genes in human B-cells contains both Top1 and Top2 activity whereas Top1 is also frequently observed in low expressed genes (Kouzine, Gupta et al. 2013). Top1 is associated with actively transcribing genes but the absence of Top1 in yeast exhibits no abnormality in transcription. In *top1Δ* mutants, Top2 fulfills the role of Top1 as both enzymes are capable of relaxing positively and negatively supercoiled DNA (Saavedra and Huberman 1986). In topoisomerase double mutants, transcription of several pol2 genes are affected, rRNA synthesis reduces by a factor of 10 and total poly$(A)^+$ RNA synthesis reduces by a factor of 3 (Brill, DiNardo et al. 1987). During transcription, the DNA positive supercoiling generated in front of the RNA polymerase would decondense the chromatin fiber and help the passage of incoming RNA polymerase (Lee and Garrard 1991).

Vertebrate cells express two genetically distinct isoforms (Top2α and Top2β) with similar structures but different biological roles. The Top2α and Top2β share a high degree

of sequence homology with 68% identity and 86% similarity (Austin, Sng et al. 1993). Moreover, they have the same capacity for complementing essential topoisomerase II functions in the absence of Top2 temperature-sensitive yeast mutants (Meczes, Marsh et al. 1997). Despite these similarities, the two isozymes play different biological roles in vertebrate cells. Human cell lines in the absence of Top2α isoform encounter serious problems at mitosis because the chromosome segregation is affected. For similar reasons, mouse embryos lacking the TOP2α gene, fail to develop beyond the 4 to 8 cell stage. In contrast, mammalian cell lines lacking Top2β pass normally through mitosis, and it is required only for aspects of nerve growth and brain development (Linka, Porter et al. 2007).

The yeast genome is small and approximately 70% of genome is made up of protein coding genes where the genes are placed very close to each other and in many cases the promoters overlap between the pair of genes. A study (Tsochatzidou, Malliarou et al. 2017) using genomic transcription run on (GRO) for in vivo labelling of nascent RNA to quantify the transcription rate and mRNA levels, showed topologically co-regulated 116 gene clusters having seven or more genes in each cluster. The gene clusters tend to be up or down regulated by Top II inactivation (*top2–4*). The down regulated gene clusters are essential genes which are placed close to pericentromeric region whereas the up regulated gene clusters are placed closed to the nuclear periphery and contains longer intergenic spaces. Upon longer transient inactivation of Top2 (>120 mins), Pol II stalls in long transcripts greater than 3KB (Joshi, Pina et al. 2012). The change in RNA abundance due to *top2* inactivation is based on the duration of transient *top2-ts* inactivation, suggesting a time dependent mechanism of stress response (Joshi, Nikolaou et al. 2018). Additionally, Top1 assists the transcription factors related to TATA box binding protein (TBP) in order to bind to TATA box and to regulate transcription initiation (Figure 2).

2.4 Topoisomerases coordination during replication

Chromosome replication is crucial and coordinated with other biological process, such as transcription, chromatin remodeling and recombination. The separation of DNA strand by helicases during replication produces helical overwinding or positive supercoiling ahead of the un-replicated portion of the replication fork (Wang 2002) (Champoux 2001) (Figure 3A). The positive supercoiling needs to be resolved for effective movement of DNA

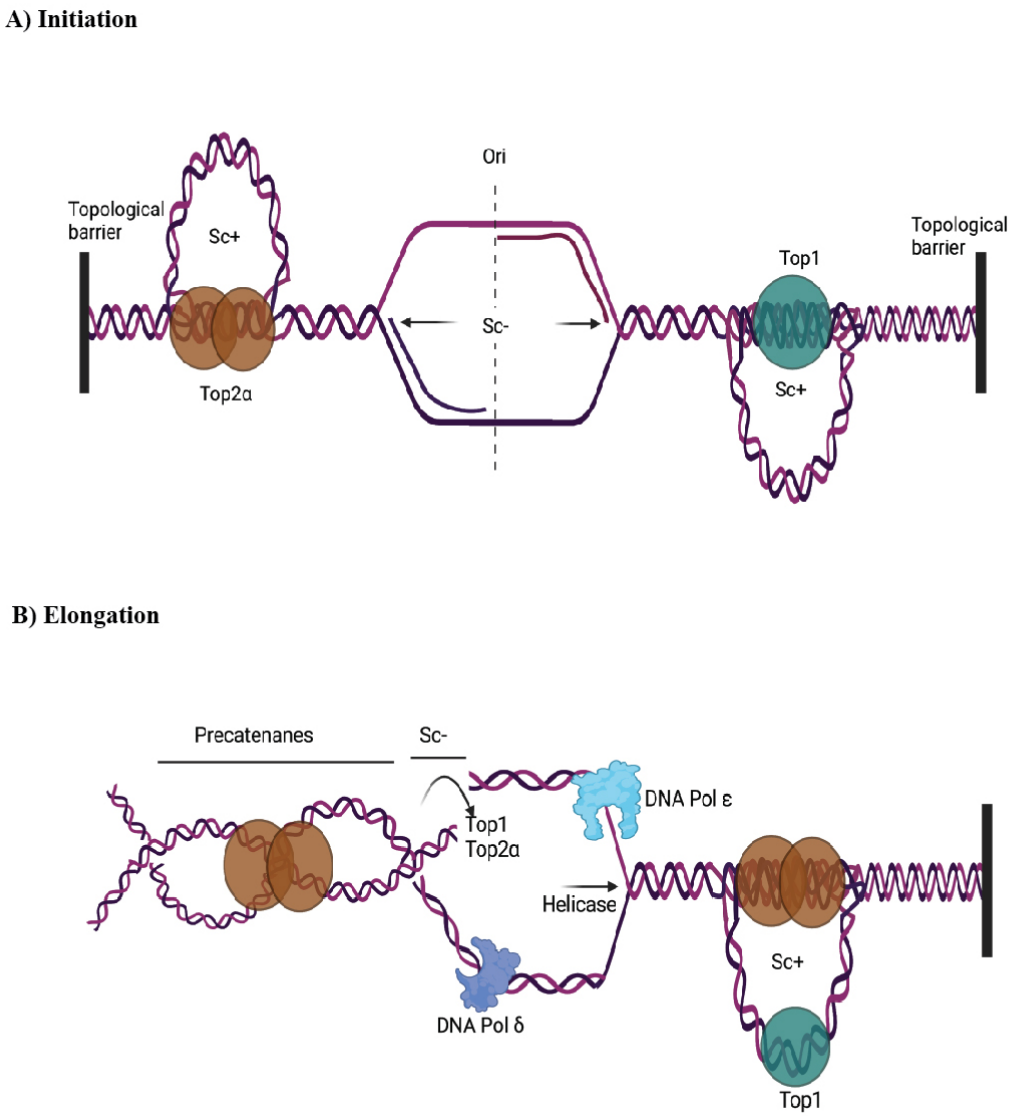


Figure 3: Topoisomerase Top1 and Top2 role in replication. A) During intiation of replication process the DNA strands are seperated which generates negative supercoil at the replication origin and positive supercoil in the flanking regions of origins, where Top1 resolves the positive supercoil and Top2α resolves the negative supercoil constraint. B) During replication elongation the positive supercoil generated ahead and negative supercoil behind the replication fork, where the supercoil and precatenanes are resolved by topoisomerases.

helicases. The progression of replication forks produces high topological constraints which are solved by topoisomerases through cleavage, strand passage and re-ligation (Figure 3B). Topoisomerases I and II result in different DNA linking number where the linking number defines the number of times a strand of DNA winds around the helical axis when the axis is constrained to lie in a plane. DNA topoisomerases catalyze the strand passage and change the linking number of DNA strand. Improper coordination between the replication fork progression and topoisomerase mediated resolution would lead to fork collapse and double strand break formation. Also, the daughter duplexes experience mechanical strain during replication by rotation of DNA at the replication fork branching point creating precatenates. Failure to resolve the precatenanes would lead to physical knots with the sister chromatids and prevent segregation during mitosis (Postow, Crisona et al. 2001).

In budding yeast, the inactivation of Top1 and Top2 affects fork integrity and activates DNA damage checkpoint. Both Top1 and Top2 associate with replication forks and additionally, Top2 accumulates in the intergenic regions at gene promoters (Bermejo, Doksani et al. 2007). Removal of Top1 doesn't affect the fork progression and also does not activate Rad53 dependent DNA damage checkpoint. Whereas the absence of Top2, does not affect the fork progression but activates Rad53 DNA damage checkpoint upon completion of mitosis (Figure 4A). Topological constraints can block the fork progression in the absence of both topoisomerases Top1 and Top2, which results in DNA breaks, fork collapse at nicks or resection of nascent chains. The exonuclease (Exo1) influences both the resection of nascent chains when there is block in the replication fork (Cotta-Ramusino, Fachinetti et al. 2005) as well as double strand break resections along with Mre11 (Takata, Tanaka et al. 2005). In the absence of both Top1 and Top2, the replication fork progression gets blocked and Rad53 checkpoint signal in S phase gets activated (Figure 4B). In *top2* mutants, aberrant S-phase events cause DNA break during cell division (Baxter and Diffley 2008). In *top2-1* cells, the Histone H2A phosphorylation on Ser129 (γH2A) which represent the DNA breaks

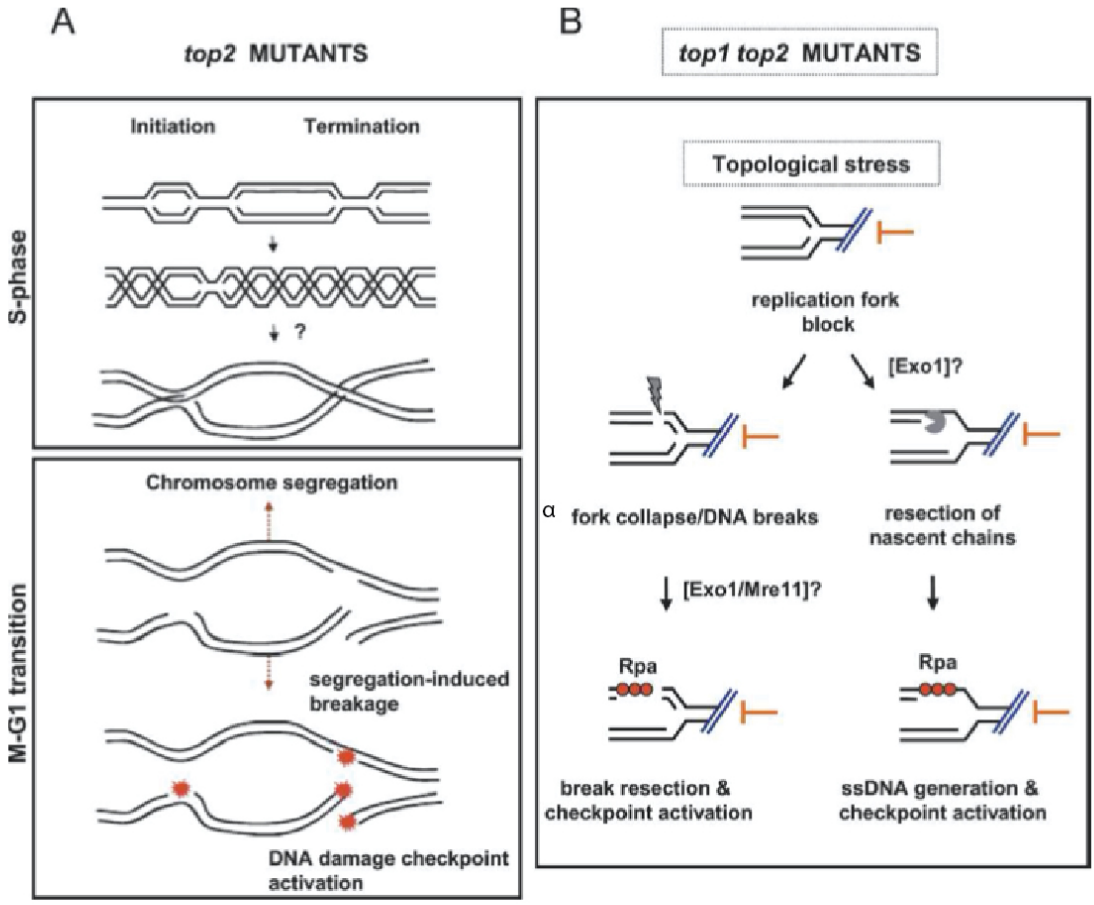


Figure 4: DNA damage and checkpoint activation in the absence of Top1 and Top2. A) In top2 mutant in S phase, precatenate forms during DNA replication initiation and termination. Also unscheduled strand passage occurs mediated by topoisomerase Top1 may lead to sister chromatid interlocking. Upon anaphase onset, tension from the mitotic spindle lead to separation of the entangled chromatids and resulting in DNA breaks and DNA damage checkpoint signals. B) In top1Δtop2-1 mutants, topological constraint cause the block of fork progression resulting in DNA breaks. The stalled replication forks could lead to fork collapse at DNA nicks, or resection of nascent chains. Exo1 is likely to remove nascent chains (Cotta-Ramusino et al. 2005) and also in double strand break removal together with Mre11 (Nakada et al. 2004). In both cases, RPA–ssDNA filaments could lead to checkpoint activation. (Figure adapted from Bermejo, Doksani et al. 2007)

accumulate significantly in Top2-bound regions in chromatin (Fachinetti, Bermejo et al. 2010). The replication termination (TER) zones where two forks converge have also more accumulation of gamma H2A that are close to the centromere region. In *top2-1* cells, the accumulation of topological constraint could lead to amplification or deletion of TER sites (Fachinetti, Bermejo et al. 2010).

The centromeres and rDNA accumulate more topological stress compared to other regions during replication. In Top2 ablated conditions, both centromeres and rDNA are linked to genome instability and sister chromatid exchange of the rDNA repeats (Christman,

Dietrich et al. 1988). The cells without Top2 in S phase causes additional DNA topological stress leading to endogenous DNA damage around centromeres due to increase cohesin activity whereas in normal condition, Top2 resolves the topological stress (Minchell, Keszthelyi et al. 2020).

2.5 Topoisomerases in secondary chromatin structure

In yeast, using DNA electron microscopy the topoisomerase type I or type II enzyme was observed at helix-helix juxtaposition (two DNA helix crossover) on negatively supercoiled plasmids containing as few as four crossovers (Zechiedrich and Osheroff 1990). The presence of Top1 and Top2 was independent of torsional stress as the enzymes were also observed at crossovers on linear DNA. The DNA helix crossovers are more prominent in supercoiled regions as compared to relaxed DNA regions. Top2 prominently binds at intergenic regions in the genome which contains low nucleosome density and in the absence of Top2, γ H2A accumulates in intergenic regions (Bermejo, Capra et al. 2009). Top2 catalyzes the strand passage of two independent double stranded DNA one after the other and it has been implicated in higher-order chromatin organization (Li, Chen et al. 1999). Intrachromosomal looping between distant regulatory elements, chromatin remodeling complexes and transcription factors helps in transcription initiation (Schneider and Grosschedl 2007). Gene looping between transcription initiator and terminator regions may facilitate polymerase recycling and to increase transcription rates (Ansari and Hampsey 2005). Top2 colocalizes with Hmo1 protein (a member of the HMGB (high mobility group box) class family). HMGB1 physically interacts with Top2 on catenated DNA structures (Stros, Bacikova et al. 2007). In the absence of Top2, Hmo1 is deleterious and accumulates at pericentromeric regions in G2/M (Bermejo, Capra et al. 2009). Together, Top2 and Hmo1 bind in intergenic regions and prevent chromosome fragility. Top2 also mediates chromosome looping (Blasquez, Sperry et al. 1989) at the transcribing regions (Figure 5A). During replication, the moving forks carry Top1 and Top2, where Top1 acts to resolve

supercoiling in front of the forks and Top2 is involved in precatenates resolution behind the fork. When fork encounters DNA loops, it leads to interlocked sister chromatid junctions (Figure 5B). Top2 then gets dislodged and Top1 catalyzes the single-strand passages at triplex junctions forming intrachromosomal catenation (Zechiedrich and Osheroff 1990) (Figure 5C). The non-B DNA structures are stabilized by Hmo1 and after replication fork passage, the specific region gets converted into sister chromatid bridges (Bianchi, Beltrame et al. 1989) (Figure 5D). In the absence of Top2, Top1 generates more complicated substrates resulting in massive DNA entanglement (Bermejo, Capra et al. 2009) (Figure 5E).

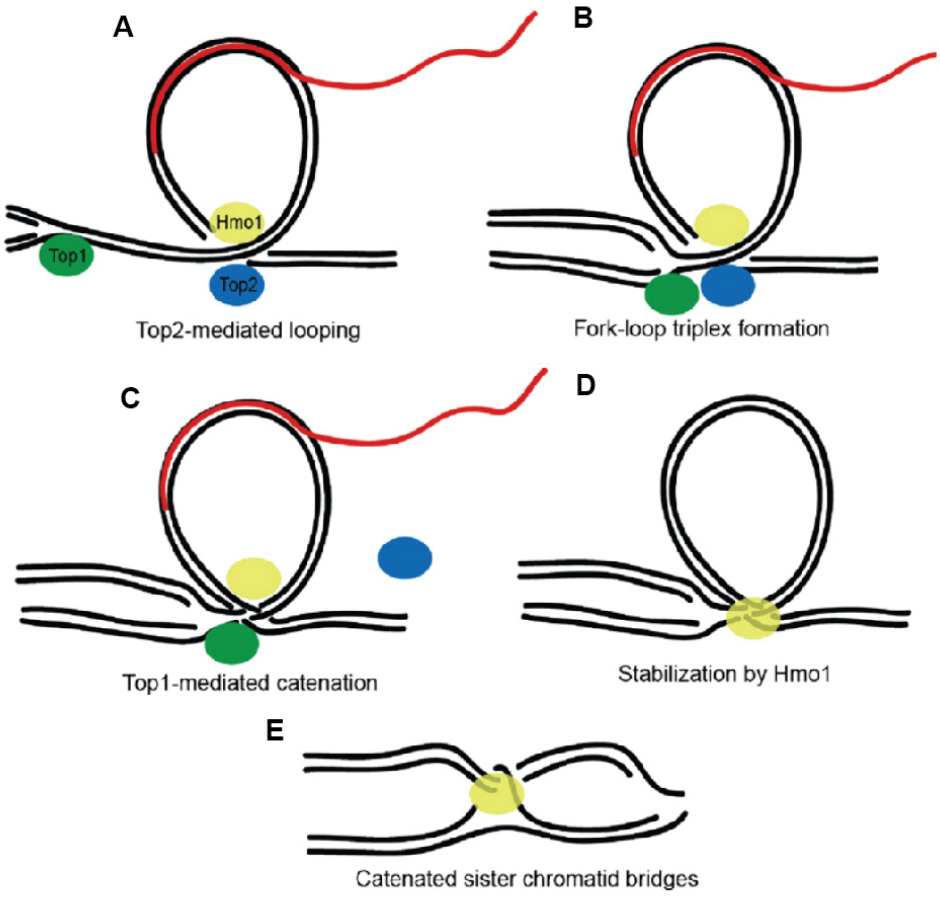


Figure 5: DNA topoisomerase role in chromatin loop formation. A) Top2 mediated DNA looping at transcribing regions. when the replication fork approaching the chromatin loop, single-strand DNA stretches together with precatenanes generating triplex structures B) Top2 might dislodge and Top1 mediate in single-strand passages at the triplex junction thus forming intrachromosomal catenation C) The precatenated structures are stabilized by Hmo1 D) and converted into sister chromatid bridges E) In the absence of Top2, Top1 could generate more substrates for Hmo1, thus leading to DNA entangling. (Figure adapted from Bermejo, Capra et al. 2009)

2.6 Chromatin organization

The chromatin fibers are diverse and are non-randomly folded in three-dimensional space. The frequency of interaction or close contact between two genomic loci of the same chromosome or different chromosome reveals the spatial organization of chromatin inside the nucleus (Dekker, Rippe et al. 2002). The study of interphase chromosome organization has shown that decondensed chromosome conformation maintains an ordered structure needed for the regulation of specific set of genes for a given tissue and also the expression of housekeeping genes in all cells (Dekker and Mirny 2016). The chromosomes occupy distinct territories inside the nucleus which can be categorized into chromosomal compartments (A/B compartments) and further sub-categorized into topologically associated domains (TADs) and chromatin loops which are mediated by specific proteins (Figure 6). In mammals, the TADs are mega-base long chromatin domains that showed increased internal contacts (Lieberman-Aiden, van Berkum et al. 2009).

Hi-C intrachromosomal contacts are used to map the TADs, manifested as triangles in the heat map, which contain increased local contact. The two distant genomic loci with frequent contacts, will be in the same TADs. A wide range of organisms exhibit TADs, having similar size, structure and architectural proteins (Mizuguchi, Fudenberg et al. 2014). The chromatin folding plays a crucial role in gene regulation, recombination, cellular development and differentiation (Dixon, Jung et al. 2015, Krijger and de Laat 2016). Perturbation of TADs structure by altering their boundaries leads to disruption of the contacts between cis-regulating elements and gene promoters, which contributes to developmental defects and cancer (Franke, Ibrahim et al. 2016) (Hnisz, Weintraub et al. 2016). In mammals, the TADs formation involves the active process of chromatin loop extrusion (Nuebler, Fudenberg et al. 2018).

Many DNA associated proteins are involved in maintaining the local structure of chromatin by folding chromosome into hierarchical domains at different genomic scales.

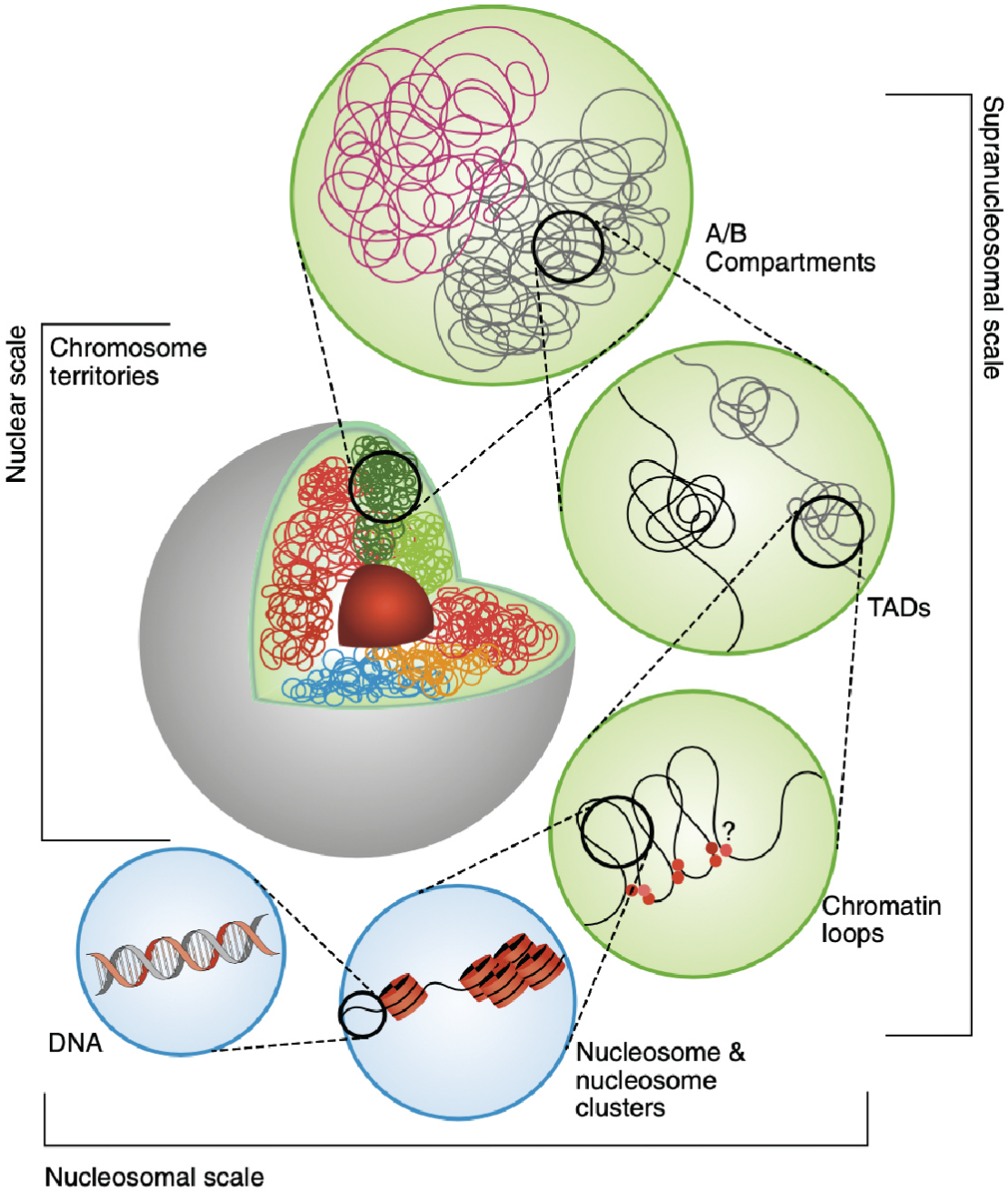


Figure 6: Schematic representation of hierarchial chromatin organization. Chromosome organized into chromosome territories where it can be further categorized into A and B compartments representive active and repressed chromatin. The compartments contains topologically associated domains (TADs) where TADs comtains increased interactions. The organization is maintained through chromatin loops mediated by architectural proteins and nucleosome formations. (Figure adapted from Dogan and Liu 2018)

The chromosomal proteins such as CTCF and cohesin localize at borders of TADs (Nora, Caccianini et al. 2020) (Gassler, Brandao et al. 2017). In the loop extrusion model, the cohesin is loaded on the chromatin fibers and these fibers pass through the cohesin rings increasing the size of the loop over time and growth of these loops are stopped when they encounter CTCF proteins thus forming a stable topological associated domain (Fudenberg,

Imakaev et al. 2016). TADs-like structures have not been reported in budding yeast (Duan, Andronescu et al. 2010). There is also no CTCF protein in yeast, but self-interacting small domains have been detected using Micro-C method (Hsieh, Weiner et al. 2015) where micrococcal nuclease is used to produce small chromatin fragments, followed by proximity ligation. These domains are in size of ± 5kb in size and the boundaries of these domains are enriched with highly expressed gene promoters, chromatin remodeling complex and cohesin loading factor Scc2.

2.7 Ligation based technique for mapping of chromatin interaction

The chromosome territories are first confirmed using fluorescence in situ hybridization (FISH) and it is shown that highly transcribing genes or active genes are present in interior of the nucleus, whereas the low transcribing genes or inactive genes are largely reside in nuclear periphery (Manuelidis 1985). In recent years, chromosome conformation capture (3C) technique has become popular and this has given more insights into the organization of chromosomes in eukaryotic cells (Duan, Andronescu et al. 2010).

The chromosome conformation capture technique is an assay technique used to capture the 3D organization based on the digestion of cross-linked DNA molecules with restriction enzymes, ligation of distal genomic regions that come into close proximity in 3D space followed by PCR or high throughput sequencing (Dekker, Rippe et al. 2002). The technique which detects interactions between two selected loci (one vs one) through ligation in combination with PCR is 3C technique (Cullen, Kladde et al. 1993). 3C technique was first used in erythroid cells, where the beta globin locus was shown to form chromatin loops between two distant regions to form active chromatin hubs that regulate transcription (Tolhuis, Palstra et al. 2002). The 3C technique is low throughput and cannot be used to detect long range interactions.

The 4C technique (Circular 3C) addressed these limitations, using primers to detect the interactions which occured between the loci of interest with unknown long distant

multiple loci (one vs all) (Simonis, Klous et al. 2006). The Chromatin interaction analysis by paired end-tag sequencing (ChIA-PET) is similar to other chromatin conformation capture techniques but includes chromatin immunoprecipitation to enrich the interactions that are mediated by specific protein of interest (Figure 7). This technique gives an enhanced resolution of inter and intra chromosomal contacts that are mediated by proteins (Li, Luo et al. 2017). All these chromatin conformation techniques help to map the contact probability of two distant loci in 3D space and it is important to note that many interactions are dynamic and not all the interactions will translate into biological function. In this study, the ChIA-PET technique was used to identify genome wide Top2 mediated loop formation.

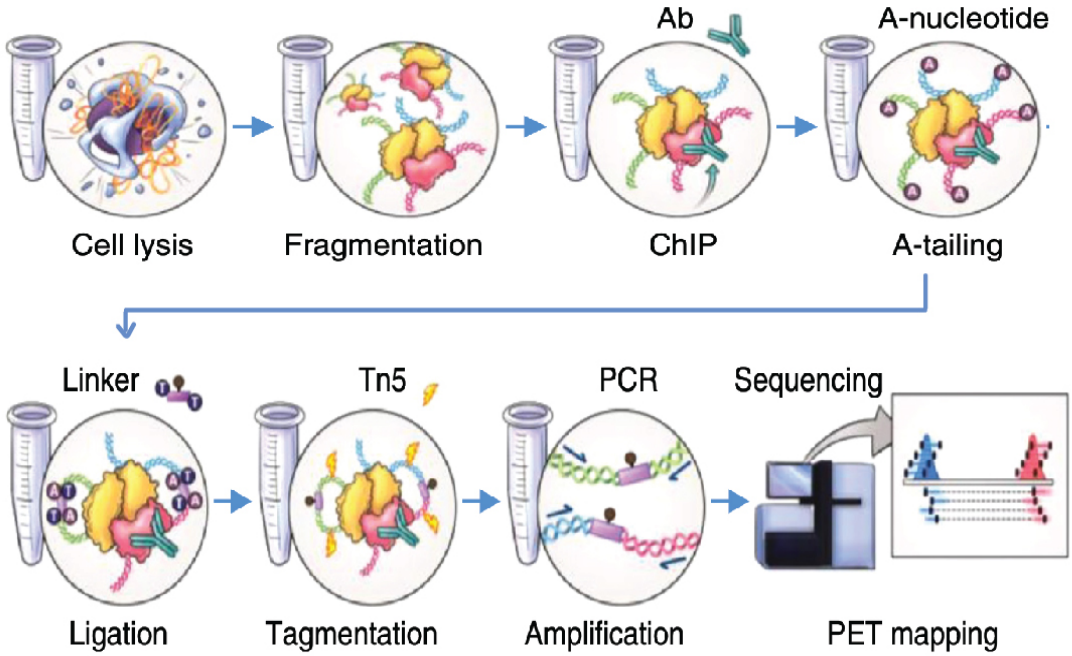


Figure 7: Schematic representation of ChIA-PET technique. ChIA-PET Protocol - Crosslinked cells are lysed and fragmented using sonication to obtain 200-300bp. Chromatin immunoprecipta-tion is performed for the protein of interest. Immunopreciptaed protein crosslinked DNA is repaired and A-tailing is performed. Proximity ligation with the bridge linker oligonucleotides containing biotin is performed. Reverse-crosslinked, Tn5 transposome digestion, ligation of sequencing adapt-ers, fragments are enriched for biotin using streptavidin beads, PCR amplification is performed and followed by paired end sequencing. (Figure adapted from Li, Luo et al. 2017)

The 4C was combined with next generation sequencing (Hi-C) to map the dynamics of chromatin contacts in various conditions like cell cycle, development and differentiation (Apostolou, Ferrari et al. 2013) (Ghavi-Helm, Klein et al. 2014). The most important

technique in chromatin organization research is Hi-C technique where the all the possible interactions are mapped across the genome (all vs all) using high throughput sequencing (Lieberman-Aiden, van Berkum et al. 2009) (Figure 8). Using the Hi-C technique the genome compartmentalization and topological associated domains are studied (Dixon, Selvaraj et al. 2012). The Hi-C interactions are visualized using 2D heatmaps where the inter and intra chromosomal contact frequencies across the whole genome are scaled to a particular resolution (bins; example 5kb, 10kb, 25kb or 50kb) and are plotted for a specific

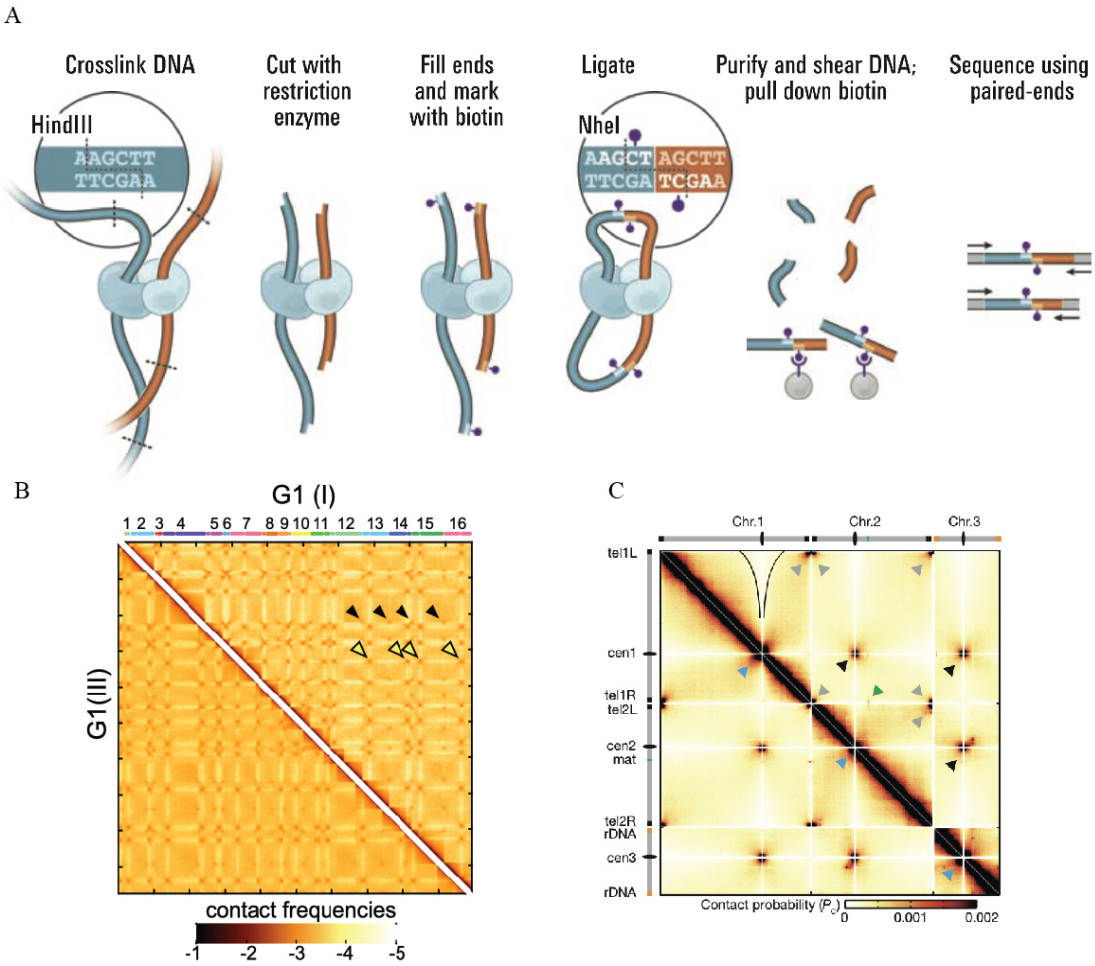


Figure 8: Schematic representation of Hi-C technique and model 2D Hi-C heatmaps. A) Schematic representation of Hi-C technique: chromatin crosslinking with formaldehyde, digestion with a restriction enzyme (RE), End filling with biotinylated nucleotide, proximity ligation, DNA purification, sonication, enrichment of biotin tagged DNA using streptavidin beads, PCR amplification and followed by paired end sequencing. (Figure adapted from Lieberman-Aiden, van Berkum et al. 2009) B) Hi-C heat map of yeast G1 phase showing centromere clustering and telomeres clustering from (Lazar-Stefanita, Scolari et al. 2017) black arrowheads indicate inter-telomere contacts and yellow arrowheads indicate inter-centromeric contacts. C) Hi-C heat map of Pombe showing centromeres clustering and telomeres clustering (Mizuguchi, Fudenberg et al. 2014) blue arrowheads indicate centromere proximal arm–arm interactions; green arrowheads indicate mat–tel1R interaction; grey arrowheads indicate intra chromosomal telomere interactions; Black arrowheads indicate interchromosomal centromere-centromere interactions.

chromosome location or across the whole chromosome. In *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe,* there are very limited inter-chromosome interactions where the centromeres interact with other centromere proximal regions through inter-chromosome interactions (Figure 8B and 8C). There are clusters of telomeres interacting with other telomeres forming a unique organization required for chromosome duplication and segregation (Lazar-Stefanita, Scolari et al. 2017) (Mizuguchi, Fudenberg et al. 2014). The organization in yeast is different from other higher eukaryotes where they have specific compartmentalization and topological associated domains.

2.8 Rabl configuration to prevent topological entanglement

Chromatin condensation during genome organization promotes topological entanglement of chromatin fibers and can inhibit basic DNA transactions like chromosome segregation and gene expression. There is a different degree of chromatin condensation that occurs in the active and inactive regions of chromosomes. There are two general types of organization, Rabl configuration where the centromeres and telomeres are at opposite poles of nucleus and the domain-based organization which has distinct chromosome territories (Figure 9). In many eukaryotes like budding yeast and *Drosophila,* chromosomes are in Rabl configuration (Berger, Cabal et al. 2008) which are characterized by clustering of centromeres on one side of the nuclear envelope and the sub-telomeric regions are positioned near the nuclear periphery. The Rabl configuration are associated with chromosome segregation and DNA repair processes.

In budding yeast, the Rabl configuration is preserved throughout the entire cell cycle (Duan, Andronescu et al. 2010). Yeast centromeres are short, approximately 125 bp sequences which are held together by microtubules from the spindle pole body at the nuclear envelope (Jin, Fuchs et al. 2000). It is shown using mathematical simulation that Rabl configuration is a key genome organizational feature which significantly reduces the topological entanglement during interphase (Pouokam, Cruz et al. 2019). Using Hi-C

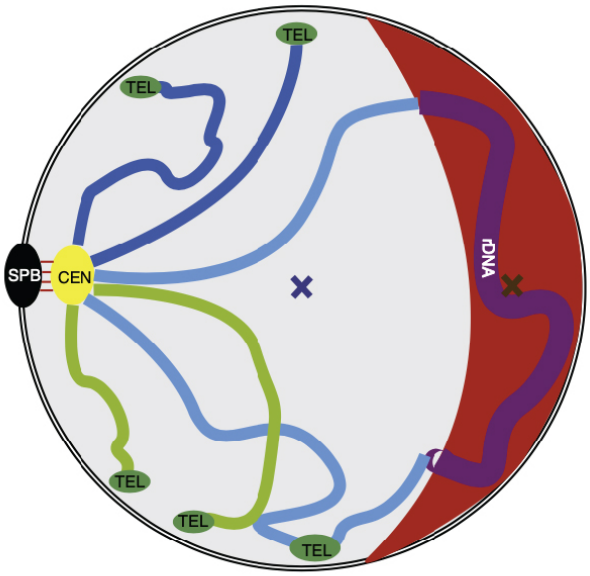


Figure 9: Model Rabl configuration of yeast genome. Centromeres (CEN) are clustered and attached to spindle pole body by microtubules (red lines). Chromosome arms are seperated and telomeres (TEL) are distributed near the nuclear envelope (NE). Nucleolus (red crescent shaped structure) contains rDNA is located opposite to spindle pole body (Figure adapted from Wang, Mozziconacci et al. 2015).

technique in yeast, the small chromosome arms make frequent contacts with other chromosomes compared to the larger chromosome arms (Duan, Andronescu et al. 2010). The chromosome III (small chromosome ~0.35MB) interacts with large segments of other chromosome arm and occupies a more crowded terrain in the 3D space of the nucleus. The chromosome IV arm (large chromosome ~1.5MB) interacts less with other chromosome arm and occupies a less crowded terrain in the 3D space of the nucleus (Duan, Andronescu et al. 2010). Another important feature is inter-chromosomal contacts between telomeres (Schober, Kalck et al. 2008) which form five to eight foci within the interphase nucleus. Yeast telomeres consist of 250-300 bp of tandem repeats. The binding of Rap1 and silent information regulator (Sir2, Sir3 and Sir4 proteins) assembles the heterochromatin at the sub telomeres and mediates transcriptional silence (Gotta, Laroche et al. 1996). In budding yeast, the long-range interactions between telomeres can be altered via the telomere-associated proteins Ku and Sir4p which are involved in anchoring yeast telomeres to the nuclear envelope. It is shown that disruption of anchorage at one end of chromosome 6 significantly reduces inter-chromosome telomere interactions (Bystricky, Laroche et al. 2005).

2.9 rDNA Organization in budding yeast

Nucleolus a discrete, crescent shaped nuclear compartment that occupies one third of the nuclear volume is located opposite to the spindle pole body (Figure 9). rDNA inside the nucleolus is formed of 100 to 200 units of 9.1 kb repeated in tandem in the middle of the right arm of chromosome 12 (Figure 10). A single rDNA unit consists of two transcribing regions (35S precursor rRNA and 5S rRNA coding regions) and two non-transcribing regions or intergenic spaces (NTS1 and NTS2). The 35S precursor rRNA and 5S rRNA regions are transcribed by RNA Polymerase I and III. The non-transcribing regions consist of origin of replication (ARS) and replication fork barrier (RFB). Due to their repetitive structure, it is highly fragile with frequent rearrangements (Kim and Wang 1989). When the large repeating rDNA unit encounters damage, it is repaired by homologous recombination

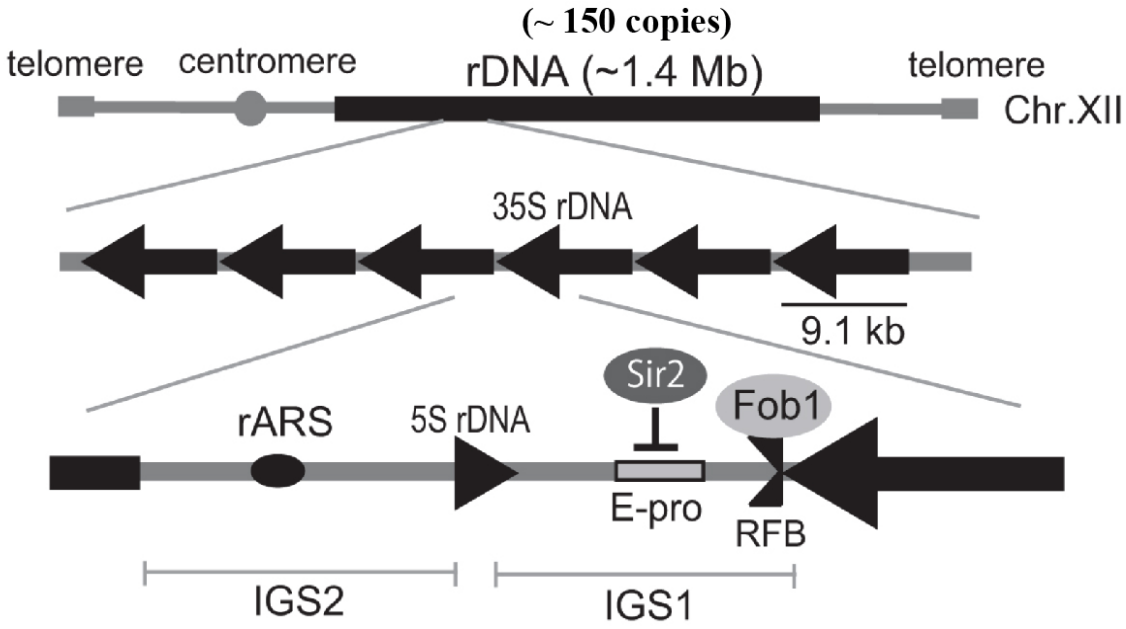


Figure 10: Yeast rDNA copies in chromosome XII. In normal condition there are approximately 150 copies of rDNA units containing total size of approximately 1.4 Mb. Each rDNA unit is 9.1kb size with two transcribing units 5S and 35S rDNA where the arrows indicates the direction of transcription. In between two rDNA units, there are two intergenic spacer (IGS1 and IGS2). IGS2 contains replication origin (rARS) and IGS1 contains replication fork blocking site (RFB) and E-pro (bidirectional promoter for noncoding transcription). Sir2 represses transcription from E-pro and Fob1 inhibits replication fork passing. (Figure adapted from Ide, Saka et al. 2013)

with a neighboring unit and as a result the number of rDNA repeat decreases (Kobayashi, Heck et al. 1998).

The rDNA copies are tightly regulated and the copy number is maintained by gene amplification. The amplification is dependent on the transcription from the noncoding bidirectional promotor E-pro (Kobayashi and Ganley 2005). The rDNA copy number loss by recombination is recovered by gene amplification which relies on stalling the replication fork by Fob1 and replication fork barrier (RFB) sequence. As result of fork stalling, double strand break occurs which is repaired by homologous recombination between sister chromatids. The double strand break repair is regulated by the histone deacetylase Sir2 and the bidirectional promoter E-pro (non-coding promoter) (Saka, Takahashi et al. 2016). When there is loss of rDNA copy number, bidirectional E-pro transcription is activated and cohesin is removed and unequal recombination occurs to increase the rDNA copy number. In case of stable rDNA copy number, the Sir2 represses the transcription from E-pro and the region is occupied by cohesin leading to equal sister-chromatid recombination in order to maintain the same number of rDNA units (Saka, Takahashi et al. 2016). Factors affecting the E-pro transcription and cohesin association are expected to alter the ratio of equal and unequal sister-chromatid recombination. In the absence of Sir2, the number of rDNA units become half compared to the wildtype (Kobayashi, Horiuchi et al. 2004) (Kobayashi and Ganley 2005).

The rDNA units replicate throughout the mitotic S phase in yeast. The actual number of rDNA copy number varies. Budding yeast contains ~150 rDNA repeats and in human there are ~300 rDNA repeats. More than 10% of yeast genes ~708 non-essential genes are involved in rDNA maintenance (Saka, Takahashi et al. 2016). Among the 708 genes involved in rDNA maintenance, 244 genes were associated with DNA repair and 142 genes were involved in genome maintenance and chromatin organization. In yeast, the rDNA stability affects the replicative life span (number of cell division throughout the life span).

In yeast cells harboring DNA topoisomerases I and II double mutants, the rDNA units are excised as extrachromosomal rings having one or more rDNA units. The expression of plasmid-borne Top1 or Top2 genes, results in the integration of rDNA rings back into the chromosome. The excision and reintegration of rDNA into the chromosome occur by recombination through homologous sequences (Kim and Wang 1989). The inactivation of either DNA topoisomerase I or II causes the mitotic recombination of rDNA, whereas the excision of rDNA repeats in the extrachromosomal ring only occurs in the double mutant (*top2-4&top1Δ*) condition. The topology of the DNA plays an important role, where the supercoil generated by transcription influences the excision or integration of rDNA units. The low level of topoisomerase II is not sufficient to relax the supercoiling generated by transcription which stimulates the rDNA intrachromosomal recombination (Kim and Wang 1989). The rDNA synthesis and maintenance inside the nucleolus is dependent on the DNA topoisomerases I and II.

2.10 Cohesin complex in budding yeast

The ring shaped cohesin complex is made of four subunits – Smc1, Smc3, Scc1 and Scc3 which are essential for holding the sister chromatids together during the replication until chromosome segregation in mitosis (Michaelis, Ciosk et al. 1997) (also see Figure 11). The Smc1 and Smc3 are the members of a family of proteins known as SMC proteins, that are responsible for the structural maintenance of chromosomes. The cohesin is a 45nm long intramolecular coiled-coil region with globular ATPase head domain on one end and a dimerization domain on the other. The Scc1 subunit bridges the two heads of Smc1-Smc3 heterodimer while Scc3 stabilizes this complex by binding to Scc1 (Haering, Lowe et al. 2002) (Figure 11A). The inner diameter of the cohesin ring is ±35nm (Anderson, Losada et al. 2002) and this is sufficient to allow the cohesin ring to move along the chromatin fiber, assuming the DNA is packed with 10nm nucleosomes and it is also feasible for the cohesin

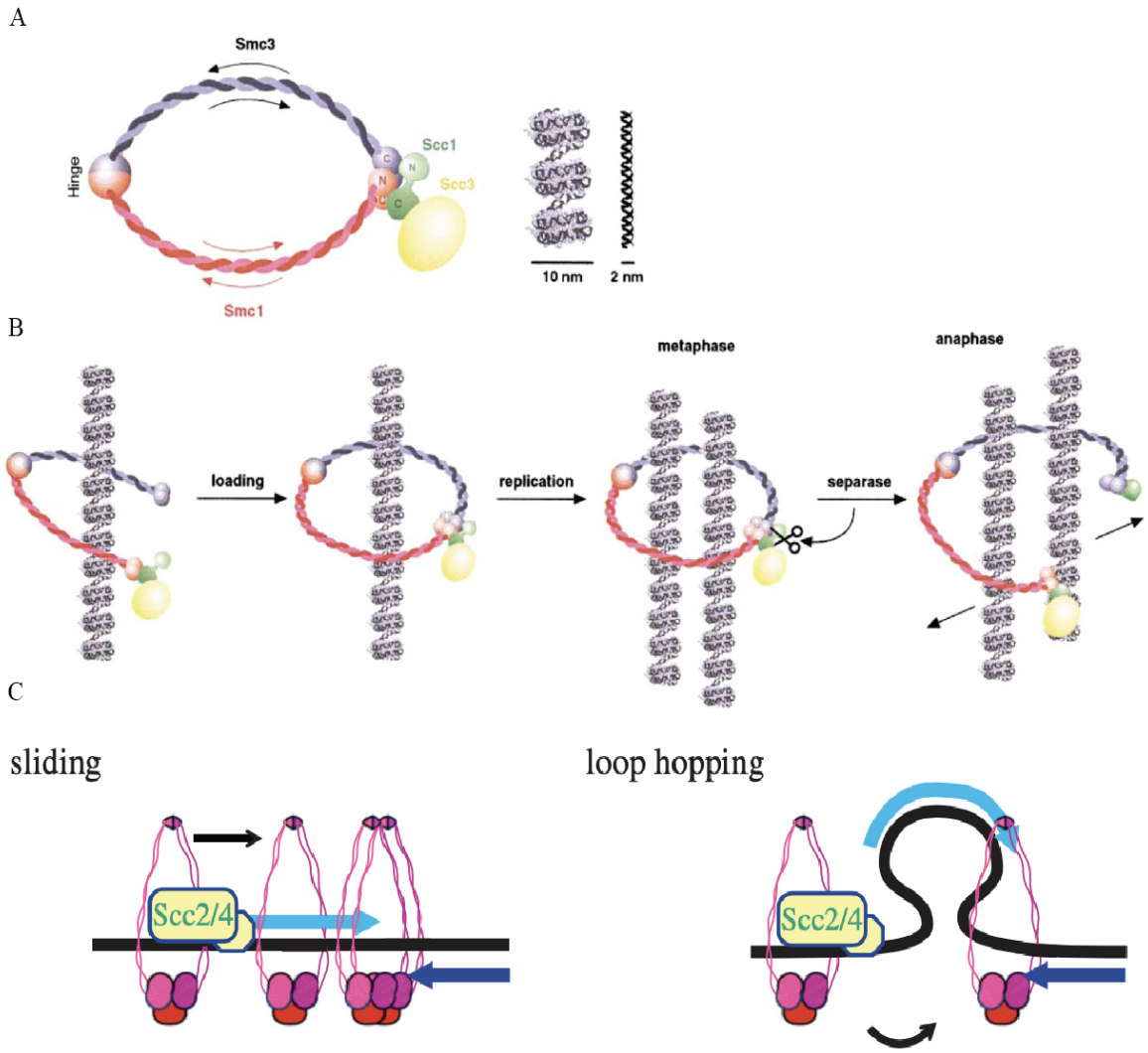


Figure 11: Model of Cohesin complex in Yeast. A) Cohesin complex contains Smc1 and Smc3 which form a hetrodimer with intramoleclar coiled coils and Scc1 bridges the Smc1 and Smc3 and links to Scc3. The 10nm chromatin fibre and 2nm naked DNA molecules are shown to scale. B) Cohesin is loaded before replication and during replication fork progression the cohesin entraps both sister chromatids. During metaphase to anaphase transition, Scc1 is cleaved and later sister chromatids can be pulled to opposite spindle poles. (Figure adapted from Haering, Lowe et al. 2002). C) Schematic representaion of cohesin translocation by sliding & loop hopping between promoter and terminator through gene looping. The blue arrows indicates the RNA pol2 transcription directions and black arrow indicates the movement of cohesin complex. (Figure adapted from Ocampo-Hafalla, Munoz et al. 2016)

to hold together the loci containing two sister chromatids, again corresponding to 10nm nucleosome fibers.

The evolutionary conserved Eco1 protein in yeast is required for regulation of sister chromatid cohesin during S phase, where it acetylates cohesin rings on their Smc3 subunits in order to stabilize cohesin's grip on chromosomes (Rolef Ben-Shahar, Heeger et al. 2008) (Ivanov, Schleiffer et al. 2002). This acetylation of Smc3 acts as a sensor mechanism for

DNA which enables the ATP hydrolysis that is required for the cohesin ring to enter or exit the DNA (Murayama and Uhlmann 2015). The acetylated cohesin prominently associates with chromosomes.

In budding yeast, the dissociation of sister chromatid cohesin is dependent on a separating protein called Esp1(separin). Esp1 cleaves the Scc1 protein subunit by proteolysis (Uhlmann, Lottspeich et al. 1999) (Figure 11B). In budding yeast, cohesin is loaded onto chromosomes in G1 phase (Glynn, Megee et al. 2004) by cohesin loader Scc2-Scc4 across different regions most prominently towards pericentromeres, centromeres and active gene promoter. After loading, cohesin translocates more permanently towards transcription termination sites (Ocampo-Hafalla, Munoz et al. 2016) through sliding (Figure 11C) or by loop hopping between promoter and terminator otherwise known as gene looping (Figure 11D). However, in many organisms including fission yeast and mammals, cohesin loader binding sites are different from the cohesin accumulation regions. Cohesin binding to the centromeres and pericentromeres facilitates the proper biorientation of sister chromatids during metaphase segregation.

There are two models proposed for holding the sister chromatids, “one ring embrace model” where the Smc3, Smc1 and Scc1/Mcd1/Rad21 form a ring like structure present during replication to hold the replicated chromosomes at pericentromeric region (Gruber, Haering et al. 2003). Second model “two ring handcuff model” where the cohesin binds to one sister chromatid and oligomerizes with another cohesin molecule bound to another sister chromatid (Zhang, Kuznetsov et al. 2008) (Campbell and Cohen-Fix 2002) (Chang, Wu et al. 2005).

In budding yeast most of the protein coding genes are transcribed during interphase of the cell cycle. Access of DNA and RNA polymerase is crucial when the chromatin is encapsulated within the cohesin ring. Transcriptional activation can lead to translocation of cohesin rings towards the convergent gene pairs termination sites (Lengronne, Katou et al. 2004). During the sliding, the topological context of the regions is preserved to keep the

region accessible for transcription or replication. In budding yeast, cohesin loaded at the centromeres follows a different mechanism compared to the cohesin loaded at other places of the genome. A study showed that upon gene activation the cohesin at non centromeric regions slides along the gene and moves to the termination site without displacement by loop hopping (Figure 11C and D). On the other hand, upon gene activation the cohesin close to the centromere undergoes displacement (Ocampo-Hafalla, Munoz et al. 2016).

2.11 Role of cohesin and topoisomerase in genome organization

The chromosome duplication and segregation involved dynamic genome reorganization which is maintained by conserved architectural proteins like cohesin and condensin. A study of budding yeast using Hi-C, elucidates the role of SMC complexes cohesin and condensin in genome organization during different cell cycle phases (Lazar-Stefanita, Scolari et al. 2017). The SMC complexes cohesin and condensin, controls the chromatin structural transitions. During S phase, the cohesin involves more in long-range intra chromosomal interactions and lesser in inter-chromosomal interaction to maintain the individualization of chromosomes. During anaphase, the mitotic chromosomes are reorganized and the condensin forms loop to bridge the centromere cluster with the rDNA loci to facilitate the chromosome segregation (Lazar-Stefanita, Scolari et al. 2017).

In *Schizosaccharomyces pombe* small regions of chromatin interact to form ‘globules’ and require cohesin (Mizuguchi, Fudenberg et al. 2014). The cohesin enriched at globule boundaries differs from cohesin enriched at sister chromatid where it requires heterochromatin. The heterochromatic pericentromeric and sub telomeric regions are enriched with cohesin. With the heterochromatin-cohesin mediated chromatin compaction at centromeres being crucial for chromosome segregation. In *Schizosaccharomyces pombe*, loss of heterochromatin leads to an increase of intra and inter chromosomal interactions resulting in disruption of chromosomal integrity (Mizuguchi, Fudenberg et al. 2014).

Using coarse grained molecular simulation technique consisting of 8 chains where each chain represents 1.6Mb large chromatin portions, the role of chromatin plectonemes (derived from Greek word: plektos meaning twisted and nema meaning thread) formed during transcription induced supercoiling have been studied. Plectonemes are extrusion of coiled loops of chromatin that can push the cohesin rings along the chromatin fibers (Racko, Benedetti et al. 2018). The cohesin rings are pushed by continuous flux of supercoiling generated by RNA polymerase where Top1 is associated (Figure 12A and 12B). The

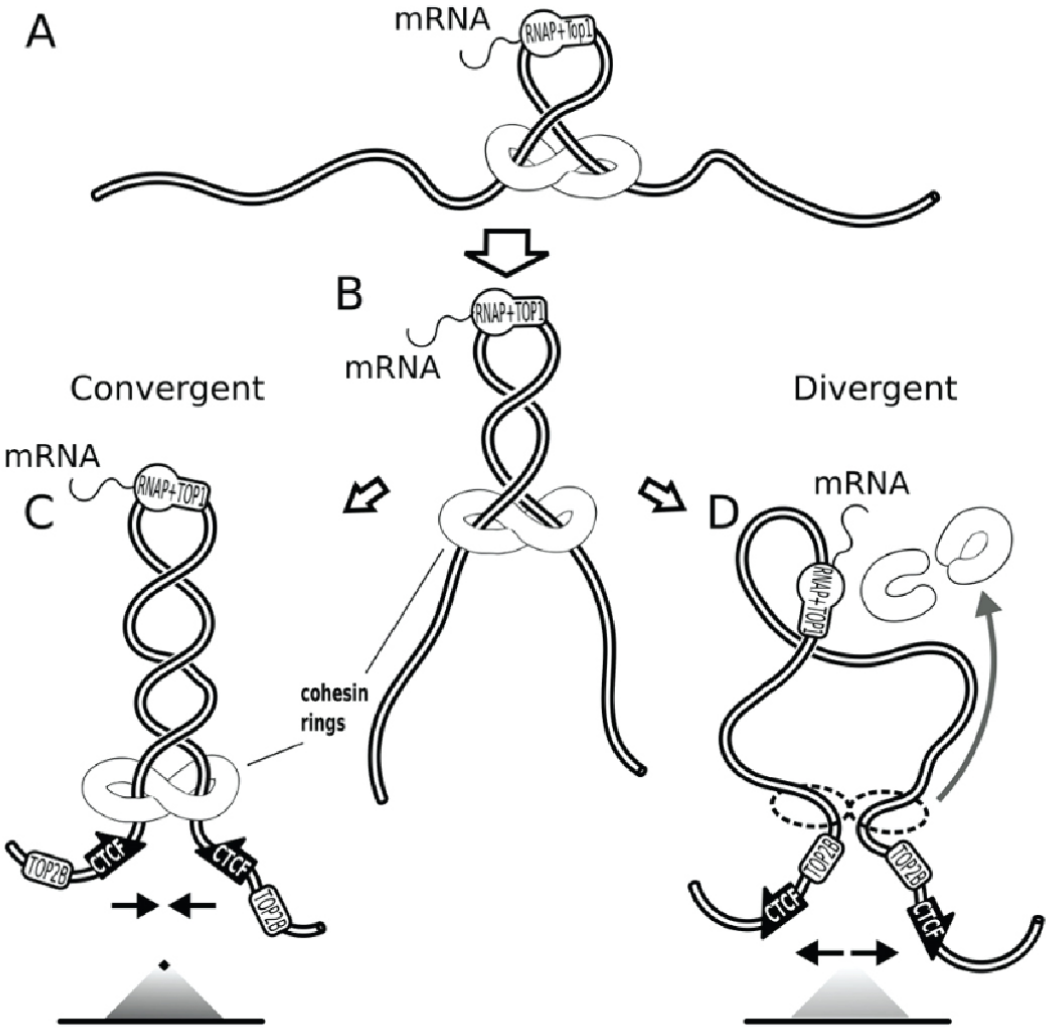


Figure 12: Cohesin sliding model based on transcription supercoiling. A) Polymerase associated with TOP1 to resolve supercoil and cohesin complex loading is enhances by negative supercoiling. B) The supercoiling based on transciption generates plectonemes which pushes cohesin handcuffs towards CTCF sites at TADs borders. C) Cohesin loops are formed when cohesin encounteres CTCF binding sites in convergent orientation. D) Cohesin dissociate from chromatin when cohesin encounters CTCF binding sites in divergent orientation this induces the relaxation of accumulated torsional stress by TOP2B. (Figure adapted from Dusan and Fabrizio et al 2017)

accumulated supercoiling from the RNA polymerase in the TADs borders are released by topoisomerase (TOPIIB) which is also located close to the TADs borders (Uuskula-Reimand, Hou et al. 2016). The cohesin loops are formed when it encounters CTCF in convergent orientation (Figure 12C), whereas cohesin loops are dissociated when it encounters CTCF in divergent orientation (Figure 12D). The model proposes that, supercoiling is the driving force of chromatin loop extrusion (Racko, Benedetti et al. 2018). There are many organisms like yeast which lack CTCF binding proteins, but there is evidence of TAD formation (Dekker and Heard 2015). The plectonemes formed due to supercoiling could act as the barrier to the movement of proteins along the chromatin fibers. At highly transcribing ORF or convergently transcribing gene pairs, high amount of supercoil accumulates to form plectonemes or non-B DNA structures that can stall large protein complexes like cohesin and help in the formation of TADs (Racko, Benedetti et al. 2018).

The aim of the thesis is (I) To study the DNA supercoiling accumulation across the yeast genome particularly in polymerase II coding genes and other functional elements like rDNA, centromere and telomeres using bTMP-CHIP technique. (II) To study the Top1 and Top2 protein accumulation across the genome and its role in maintaining the DNA supercoil secondary structures (III) The role of Top2 in chromatin loop formation using ChIA-PET technique (IV) Ablation of negative supercoil by expressing *E.coli* TopA in topoisomerases I and II double mutants and its impact in nucleosome binding using Histone H3 ChIP-sequencing technique, cohesin localization using protein ChIP-sequencing technique and higher order genome organization using Hi-C technique.

3. Materials & Methods

3.1 Strains and growing conditions

All S. cerevisiae strains are W303 derivatives. Strains were grown at 28 °C YPD medium. G1 synchronization was carried out using 3-5 µg/ml of alpha factor. For S-phase samples, G1 cells were washed twice in YP medium and allowed to grow for 15 min in fresh medium. For temperature-sensitive strains, cells were allowed to grow for 10 min in fresh medium after G1 release, centrifuged and then dissolved in pre-warmed media at 37 °C and allowed to grow for 15 min. Cell cycle progression into S phase was monitored by FACS and budding profiles. For E. coli TopA expression, wild type and top1Δ top2-1 mutants harboring either control or TopA expression plasmids were grown at 25 °C in synthetic medium (SC) lacking leucine. Cells were shifted to 37 °C for inactivation of Top2 after reaching 8×10^6 cells/ml concentration.

3.2 Strains used in this study

	Strain	**Stock Number**	**Genotype**	**Reference**
1	Wt	SY2080	Mata, ade2-1, ura3-1, trp1-1, leu2-3, leu2-112, his3-11, his3-15, can1-100, GAL, PSI+, RAD5+	Lab collection
2	*top2-1*	CY8423	MATa ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3,	Lab collection

			12 trp1-1, RAD5+, top2-1	
5	*top1⊿*	CY9950	MATa ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3, 12 trp1-1, RAD5+, top1::HIS	Lab collection
6	Top1-6XHis-10xFlag	CY7178	Mata, ade2-1, ura3-1, trp1-1, leu2-3, leu2-112, his3-11, his3-15, can1-100, GAL, PSI+, RAD5+, ura3::URA3/GPD-TK(7X), top1-6His10Flag (KANr)	Lab collection
7	*top2-1*,Top1-6XHis-10xFlag	CY7411	Mata, ade2-1, ura3-1, trp1-1, leu2-3, leu2-112, his3-11, his3-15, can1-100, GAL, PSI+, RAD5+, ura3::URA3/GPD-TK(7X), top1-	Lab collection

			6His10Flag (KANr), top2-1	
8	*top2-1top1Δ*	CY10344	MATa ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3, 12 trp1-1, RAD5+, top2-1, top1::HIS	Lab collection
12	Wt [control]	CY15421	Mata, ade2-1, ura3-1, trp1-1, leu2-3, leu2-112, his3-11, his3-15, can1-100, GAL, PSI+, RAD5+ [pYEp13-LEU empty]	This Study
13	Wt [TopA]	CY15422	Mata, ade2-1, ura3-1, trp1-1, leu2-3, leu2-112, his3-11, his3-15, can1-100, GAL, PSI+, RAD5+ [pJRW13-YEptopA-pGPD-LEU]	This Study
14	*top2-1top1Δ* [control]	CY15423	MATa ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3, 12 trp1-1, RAD5+,	This Study

			top2-1, top1::HIS [pYEp13-LEU empty]	
15	*top2-1top1Δ* [TopA]	CY15424	MATa ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3, 12 trp1-1, RAD5+, top2-1, top1::HIS [pJRW13-YEptopA-pGPD-LEU]	This Study
16	Wt [control]; Scc1-6XHis-10xFlag	CY16231	Mata, ade2-1, ura3-1, trp1-1, leu2-3, leu2-112, his3-11, his3-15, can1-100, GAL, PSI+, RAD5+, SCC1::SCC1-10X Flag-KanMX6, [pYEp13-LEU empty]	This Study
17	Wt [TopA]]; Scc1-6XHis-10xFlag	CY16232	Mata, ade2-1, ura3-1, trp1-1, leu2-3, leu2-112, his3-11, his3-15,	This Study

			can1-100, GAL, PSI+, RAD5+, SCC1::SCC1-10X Flag-KanMX6, [pJRW13-YEptopA-pGPD-LEU]	
18	*top2-1top1Δ* [control]]; Scc1-6XHis-10xFlag	CY16233	MATa ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3, 12 trp1-1, RAD5+ top1::HIS, top2-1, SCC1::SCC1-10X Flag-KanMX6, [pYEp13-LEU empty]	This Study
19	*top2-1top1Δ* [TopA]]; Scc1-6XHis-10xFlag	CY16234	MATa ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3, 12 trp1-1, RAD5+ top1::HIS, top2-1, SCC1::SCC1-10X Flag-KanMX6, [pJRW13-YEptopA-pGPD-LEU]	This Study

3.3 Growth media and buffer composition

Media and Buffer	Composition
YPD	1% Yeast extract, 2% glucose
YPD agar	1% Yeast extract, 2% glucose, 2% agar
SC	0.67% Yeast nitrogen base (YNB, DIFCO w/o Amino Acids), 2% glucose and required amino acids
TE	Tris-HCL 10mM (pH 8), EDTA 1mM (filter with 0.2µM)
PBS	137mM NaCl, 10mM PO_4 (pH 7.4), 2.7mM KCl (filter with 0.2µM)
TBS	20mM Tris-HCL (pH 7.5), 150 mM NaCl (filter with 0.2µM)
2X Reducing Laemli Buffer	4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromphenol blue, 0.125 M Tris HCl (pH 6.8)
PBS/BSA	1 X Phosphate Buffered Saline containing 5mg/ml Bovine Serum Albumin (filter with 0.2µM)
Lysis Buffer	50mM Hepes-KOH (pH 7.5), NaCl 140mM, EDTA 1mM, Triton-X100 1% Na-deoxycholate 0.1%
Lysis Buffer++	Lysis Buffer (15ml), 1 tablet Protease inhibiter cocktail tablet, 1 mM Phenylmethylsulfonyl fluoride (PMSF)

ChIP Wash Buffer 3	Tris-HCl (pH 8.0) 10mM, LiCl 250mM, NP-40 0.5%, Na-deoxycholate, EDTA 1mM
Elution Buffer	Tris-HCl (pH 8.0) 50mM, EDTA 10mM, SDS 1% (filter with 0.2µM)
TE SDS	Tris-HCL (pH 8.0) 10mM, EDTA 1mM, SDS 1%
Proteinase K	50 mg/ml in 50% Glycerol
RNase A	10 mg/ml in 50% Glycerol
TSE 1	20mM Tris-HCl pH 8.1, 2mM EDTA, 150mM NaCl, 1% Triton, 0.1% SDS (filter with 0.2µM)
TSE 2	20mM Tris-HCl pH 8.1, 2mM EDTA, 500mM NaCl, 1% Triton, 0.1% SDS (filter with 0.2µM)
TSE 3	10mM Tris-HCl pH 8.1, 1mM EDTA, 0.25M LiCl, 1% NP40, 1% DOC
Elution Buffer 2	Formamide (95%) and 10mM EDTA
Elution Buffer 3	Tris-HCl (pH 8.0) 50mM, EDTA 10mM, SDS 0.1% (filter with 0.2µM)
DNAse reaction mix	ddH_2O 14.8 µl, 10X One-Phor All Buffer plus 2 µl, 25mM $CoCl_2$ 1.2 µl, DNAse I (1U/ µl) 2 µl
Glycine	2.5 M Glycine
ChIP Wash Buffer 2	Lysis Buffer, 360mM NaCl

T4 DNA master mix	1X Buffer for T4 DNA polymerase, 100 µM 10mM dNTPs, 0.4 U/µl T4 DNA polymerase
Klenow (3′ > 5′exo-) master mix	1X NEB Buffer 2, 100 µM dATP, 0.2 U/µl Klenow (3′ > 5′exo-)
T4 DNA ligase master mix	1X T4 DNA ligase buffer, 0.57 ng/µl bridge linker (200 ng/µl), 0.02 U/µl T4 DNA ligase
Tagmentation mix	1X Tagmentation buffer, 8 µl Tagmentation enzyme from Nextera DNA Sample Prep Kit
ChIP Wash Buffer 4	5mM Tris–HCl (1 M, pH 8.0), 0.5mM EDTA (0.5 M) and 0.5M NaCl (5 M)
ChIP Wash Buffer 5	10mM Tris–HCl (1 M, pH 8.0), 1mM EDTA (0.5 M) and 2M NaCl (5 M)
iBlock buffer	2 % I-Block Protein-Based Blocking Reagent dissolved in 65 °C water bath, 5 ml of 10% (wt/vol) SDS
Hi-C Lysis Buffer	1x TBS, Protease inhibitor cocktail tablet (1 tablet for 15 ml buffer), 1% Triton X-100
Digestion buffer	1XDpnII buffer, 1XCut Smart buffer

3.4 Reagents and Instruments

Branson Sonifier 2508, Danbury, CT

Glasss beads (Sigma G-8772)
PowerLyzer 24 Homogenizer Cat No. /ID: 13155
Qiagen PCR purification kit Cat No. 28106
Zymo Research DNA Clean & Concentrator Cat No. D4004
Covaris Sonicaation E220 evolution
milliTUBE 1ml AFA Fiber Part No. 520130
microTUBE AFA Fiber Pre-Slit Snap-Cap 130 µl Part. No. 520045
Agilent 2100 Bioanalyzer Instrument
NanoDrop™ 2000/2000c
Qubit 4 Fluorometer Cat No. Q33238
Qubit™ 1X dsDNA HS Assay Kit Cat No. Q33230
GenomePlex® Complete Whole Genome Amplification (WGA) Kit
DNAse I (Gibco BRL Amplification Grade 18068-015)
10X One-Phor-All Buffer plus (Pharmacia, Piscataway, NJ 27-0901-02)
Terminal transferase (Roche, Indianapolis, IN, 222574)
Biotin-11-ddATP (NEN NEL548)
Dynabeads™ MyOne™ Streptavidin C1 Cat No. 65002
Dynabeads™ Protein G for Immunoprecipitation Cat No. 10004D
Affymetrix GeneChip Scanner 3000 7G
Affymetrix GeneChip S. cerevisiae Tiling 1.0R Array (Sc03b_MR) Cat No. 900645
Illumina Nextseq 550 System Next Generation Sequencer
NextSeq 500/550 High Output Kit v2.5 (150 Cycles) Cat No. 20024907
37% formaldehyde solution Sigma Cat No. 47608
Monoclonal ANTI-FLAG® M2 antibody mouse Cat No. F1804-1mg
anti-histone H3 antibody (Abcam, cat. no. ab1791)
dATP solution (100 mM; NEB, cat. no. N0440S)

T4 DNA ligase (NEB Cat. No. M0202M)
T4 DNA Polymerase (Promega, cat. no. M4215)
10mM dNTPs (Life Technologies, cat. no. 18427-088)
AmPure XP beads (60 ml; Beckman, cat. no. A63881)
Klenow Fragment (3′>5′ exo-) (NEB, cat. no. M0210M)
NEB Buffer 2 (NEB, cat. no. B7002S)
Nextera DNA Sample Prep Kit (Illumina, cat. no. FC-121-1031)
Protease Inhibitor Cocktail EDTA-free
Phenol:Chloroform:Isoamyl Alcohol 25:24:1, Saturated with 10mM Tris, pH 8.0, 1mM EDTA Cat No. P3803
MaXtract High Density Cat No. 129073
DpnII (NEB Cat. No. R0543L)
NEBNext Ultra II End Repair/dA-Tailing Module

3.5 Software and tools

rMAT	http://www.bioconductor.org/packages//2.11/bioc/html/rMAT.html
FASTX	http://hannonlab.cshl.edu/fastx_toolkit/
PICARD	https://broadinstitute.github.io/picard/
Bowtie2	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
DANPOS	https://sites.google.com/site/danposdoc/
MACS2	https://pypi.org/project/MACS2/
WashU Epigenome Browser:	https://epigenomegateway.wustl.edu/

Bedtools	https://bedtools.readthedocs.io/en/latest/
Cutadapt	https://cutadapt.readthedocs.io/en/stable/
Bwa	http://bio-bwa.sourceforge.net/
HiCUP	https://www.bioinformatics.babraham.ac.uk/projects/hicup/
Juicer	https://github.com/aidenlab/juicer
Juicebox	https://github.com/aidenlab/Juicebox/wiki/Download

3.6 bTMP ChIP

This method is adapted from (Naughton, Avlonitis et al. 2013) and modified for yeast genome. After reaching the concentration of 1×10^7 cells/ml in 800ml of culture (400ml culture per replicate), Sodium azide (0.1%) was used to metabolically arrest cells and to ensure the preservation of the most prevalent topological context present at any given genomic position. The cells were kept in ice for 20 mins after sodium azide treatment. We note that this method does not aim to study the dynamic topological transitions. The cells were collected by centrifugation and the pellets were treated with 800µl of Polyethylene glycol (PEG 50%), 100µl of lithium acetate (1M) and 100µl of Dimethyl sulfoxide. Permeabilized yeast cells were incubated with bTMP (400 µg per $2x10^9$ cells) in dark for 90 min and then cross-linked by 365 nm UV light at 2000 energy (millijoules/cm^2) for 4 times to form adducts between two DNA strands with psoralene. Cells were collected and washed with 1 X Milli-Q water. Cells were washed with 1ml of Lysis buffer++ and 400 µl of Lysis Buffer++ added to the cell pellet followed by glass beads. Cells were lysed with PowerLyzer 24 Homogenizer at 4000 rpm for 20 sec for 8 rounds. In between each round the cells were kept in ice for 5 mins. The sample is collected from glass beads and centrifuged to collect the pellet at 13400 RCF for 10 mins at 4°C. 450µl of Lysis Buffer ++ is added to the pellet for sonication. Branson sonicator 2508 is used to sonicate the samples for 6 times for 15 seconds by placing the samples in ice. The supernatant is collected by centrifuging at 16000

RCF for 5min at 4°C. Sonicated samples are incubated with 30 µl of 20% SDS, ProteinaseK 25 µl and RNAseA 5 µl incubated for overnight at 50°C. Samples are purified with Qiaquick PCR purification kit 8 tubes per sample (800ml culture) and eluted with 200 µl of elution buffer from the kit. The concentration of the DNA is measured using nano drop. Around 100 µg of DNA is obtained for 800ml culture. Input DNA is collected from the purified sheared chromatin (1/100 of the material is collected as Input) and stored at -20°C. Purified DNA was incubated with Dynabeads MyOne streptavidin overnight at 4°C. The beads with the samples are washed with 2 times with TSE1 buffer, TSE2 buffer and TSE3 buffer using magnetic rack. The samples are eluted with 50 µl Elution Buffer 2 in shaker for 20 mins at 90°C. 200 µl of eluted samples are used for each replicate (IP). The IP and Input samples are purified with DNA clean and concentration kit from Zymoresearch and eluted with 13 µl of elution buffer from the kit.

For bTMP-chip with naked DNA, genomic DNA was isolated from Qiagen Genomic-tip 100/G & Genomic DNA buffer Set Purified DNA was sonicated and bTMP was added to purified DNA and incubated in dark for 90 min and crosslinked with UV at 365 nm (800 mJ/cm2). DNA was precipitated using isopropanol and washed with 70% ethanol. Dried pellet was dissolved in Elution Buffer 3 and incubated with Dynabeads MyOne streptavidin overnight at 4°C. Input DNA was isolated from sheared chromatin input (1/100 of the material is collected as Input). Both IP and Input samples are processed as described in the “Microarray and data processing” section.

3.7 Microarray and data processing

Both IP and Input DNA was amplified using GenomePlex® Complete Whole Genome Amplification Kit (protocol as specified by the provider). Samples are measured in nanodrop and 5000 ng of Input and IP is used for the further steps. Biotin labelling is performed using 4.85 µl 10X One-Phor-All Buffer plus, 25 mM $CoCl_2$ 2.9 µl, DNAase reaction mix 1.5 µl and 5 µg of DNA (IP or Input) with ddH_2O in 40.75 µl. Samples are

vortexed, pulse-spun and incubated in thermocycler at 37°C for 30 seconds and then transfered to 95°C for 15 minutes. Samples are transferred to new 1.5ml Eppendorf tube and DNA labeling is performed using 5 µl of TdT reaction buffer, 1 µl Biotin-N11-ddATP (1nMole/ µl) and 1 µl terminal transferase (400 U/ µl). Samples are incubated at 37°C for 1hr. Labelled samples are hybridized to Affymetrix GeneChip S. cerevisiae Tiling 1.0R Array (Sc03b_MR) according to the Affymetrix standard protocol. The CEL files were processed using rMAT (Droit, Cheung et al. 2010) R package to identify enriched regions across the genome. At first, systematic biases such as probe effect were corrected by normalization. Then probe intensities were smoothed and a score was calculated for each probe using IP and Input. To detect enriched regions based on the probe score the following parameters were used; dMax=300 (sliding window side), nProbesMin=8 (Minimum number of probes to average), method=Score (calling enriched regions based on sliding widow scores), log2 threshold=1.5 (equal and greater than 1.5 are labelled as enriched region). For bTMP experiments, bTMP binding 'in cells' (IP/Input) were subtracted by 'naked genomic DNA' score (IP/Input) to correct for false positives binding of bTMP. Correction for microarray readings was done by subtracting bTMP binding in cells with bTMP bound to purified-sonicated DNA, that is (bTMPcells-IP/input) - (bTMPpurified DNA-IP/input), which will give the normalized ratio of bTMP (bTMP-IP/input) binding.

3.8 Protein and Histone H3 ChIP Sequencing:

Chip analysis for proteins was carried out as described previously (Bermejo, Katou et al. 2009) with a few modifications. The protein of interest is epitope tagged at C-terminus with 10X-Flag and 6X-PK for the antibody recognition. After reaching the concentration of 1×10^7 cells/ml in 200ml of culture (100ml culture per replicate), cells were crosslinked with 1% formaldehyde in culture medium for 30 min at room temperature in shaking condition followed by quenching with 0.125 M glycine for 5 mins shaking and transferred to ice for 20 mins. Cells were washed with 1ml of Lysis buffer++ and 400 µl of Lysis Buffer++ added

to the cell pellet followed by glass beads. Cells were lysed with PowerLyzer 24 Homogenizer at 4000 rpm for 20 sec for 8 rounds. In between each round the cells were kept in ice for 5 mins. The sample is collected from glass beads and centrifuged to collect the pellet at 13400 RCF for 10 mins at 4°C. 950µl of Lysis Buffer ++ is added to the pellet for sonication. The samples are sonicated using Covaris E220 sonicator with milliTUBE 1ml AFA Fiber (Parameters: Duty Factor 6, burst/cycle 200, Peak Watt 200, Time 2400 seconds). The samples are centrifuged at 16000 RCF for 5mins at 4°C and supernatant is collected. After sonication Input DNA is collected (1/100 of the material is collected as Input) and 10 µl for western blot is collected and stored at -20°C. The 100 µl of Dynabeads™ Protein G for 200ml culture is washed twice with 1 ml of PBS/BSA using magnetic rack. For proteins tagged with 10X-Flag 25 µl of 1mg anti-Flag antibody (M2-antiflag, Sigma) or For Histone H3 25 µl anti-histone H3 antibody and 75 µl of PBS/BSA is added to magnetic beads and incubated for 4 hours in slow rotation at 4°C. The chromatin fraction was incubated (four tubes per sample) with magnetic beads coated with anti-Flag antibody (M2-antiflag, Sigma) overnight at 4°C for Immunoprecipitation (IP). 5 µl of supernatant (Flow) is collected after magnetic bead precipitation for western blot and stored at -20°C. The samples are washed placing the magnetic rack in ice twice with Lysis buffer, ChIP wash buffer 2, ChIP wash buffer 3 and once with 1X TE buffer (all buffers placed in ice). Beads are centrifuged at 800 g for 3mins at 4°C. Tubes are placed in the magnetic rack and remaining liquid are removed using the vacuum pump. 40 µl of Elution Buffer is added to each tube and incubated at 65°C for 20mins shaking. The tubes are centrifuged for 1min at 16000 g at room temperature and samples were collected using the magnetic rack. 5 µl of IP is collected for western blot and stored at -20°C. 75 µl of samples (IP) per replicate is collected. IP and Input samples are treated with 95 µl of TE SDS, 10 µl of Proteinase K and 6 µl of RNaseA. Reverse crosslinking was carried out at 65 °C overnight. The western blot is performed with the IP, Flow and Input containing 2x Laemli Buffer. For good immunoprecipitation, an explicit band at the size of the protein of interest is observed for the IP fraction by western blot. The

IP and Input samples are purified with DNA clean and concentration kit from Zymoresearch and eluted with 13 µl of elution buffer from the kit. The input samples are analyzed in Agilent bioanalyzer for the optimum fragmentation. The Input and IP samples were ligated with illumina barcodes and amplified using Kapa library amplification kit, followed by size selection with AMPure XP Bead. ASPRI cleanup with a 1.5× AMPure XP Bead : DNA ratio was performed and final libraries were eluted and sequenced using Illumina Nextseq 550 System with NextSeq 500/550 High or medium Output Kit v2.5 where each sample (IP and Input) contains approximately 10 million paired end reads.

3.9 ChIP-seq data analysis

The paired end raw reads were filtered based on the basis of quality value (-q 20 and -p 30) using the FASTX Toolkit. The filtered reads were aligned to the reference genome (SacCer 2011) using bowtie2 aligner to produce alignment file (BAM). The PCR duplicates were removed from the aligned BAM files using PICARD tools. The BAM files were sorted and indexed for the peak calling using SAMtools. The bedgraph files were generated by comparing BAM files of IP and input (IP read coverage/input read coverage) resulting in a ratio for every base across the whole genome using bamCompare from deepTools.

For Protein-ChIP sequencing MACS2 peak calling tools is used to produce the bed files having enriched peak information with the following parameters (--gsize=1.21e+7 -p 0.01 --nomodel --extsize 200 –broad). For Histone H3 Peak calling was performed using the DANPOS (dpos) toolkit54 with the IP/input threshold 1.4 (−q 1.4) where the output peaks correspond to the individual nucleosome. The DANPOS was preferred over the MACS toolkit for the dynamic nucleosome analysis at single-nucleotide resolution. The bed and bedgraph files are visualized using WashU Epigenome browser and also by custom made R and python scripts.

3.10 Meta-analysis of protein coding and other genomic features

The bed and bedgraph files of bTMP ChIP, Protein ChIP sequencing and Hinstone H3 sequencing are used to analyze the binding profile of specific genomic features such as protein coding genes, centromeres, tRNA and telomeres etc. The meta-gene count and average intensity plot are used to examine the averaged enriched peak profile for a specific group of genes or for all protein-coding genes (6706 genes) with 500 bases upstream from the transcription start site and 500 bases downstream from the transcription termination site in the yeast genome. Peak scores of the bed and bedgraph files were mapped using bedtools (Quinlan and Hall 2010) for each base of the gene including upstream and downstream bases. The varying length of the gene (ORF) was scaled to 1000 bases. For scaling, the following equation was used for each base in the ORFs ((Z-x_i)/(y_i-x_i)) *1000; x_i = start position of the i-th gene i = (1,2,3 ... total genes), y_i = end position of the i-th gene i = (1,2,3 ... total genes), Z = base position (1,2,3....y_i-x_i). For the representation of average intensity, the IP/input values of the normalized position of each gene were aggregated using the median. For average gene density, the IP/input ratio was converted to either 1 or 0 (categorical) based on the threshold of 1.5 (=1.5 is 1 and 1.5 is 0) of all normalized positions (1000 bases in ORF, 500 bases upstream, and 500 bases downstream) of each gene and finally aggregated using the sum function to obtain the average total. For visualization, the scores were smoothed using the generalized additive model (GAM) to obtain a smooth curve.

3.11 ChIA-PET (Chromatin interaction analysis by paired-end tag sequencing)

This method is adopted from (Li, Luo et al. 2017) and modified for yeast genome. After reaching the concentration of 1×10^7 cells/ml in 200ml of culture (100ml culture per replicate), cells were crosslinked with 1% formaldehyde in culture medium for 30 min at

room temperature in shaking conditions followed by quenching with 0.125 M glycine for 5 mins in shaking and transferred to ice for 20 mins. Cells were washed with 1ml of Lysis buffer++ and 400 µl of Lysis Buffer++ added to the cell pellet followed by glass beads. Cells were lysed with PowerLyzer 24 Homogenizer at 4000 rpm for 20 sec for 8 rounds. In between each round the cells were kept in ice for 5 mins. The sample is collected from glass beads and centrifuged to collect the pellet at 13400 RCF for 10 mins at 4°C. 950µl of Lysis Buffer ++ is added to the pellet for sonication. The samples are sonicated using Covaris E220 sonicator with milliTUBE 1ml AFA Fiber (Parameters: Duty Factor 6, burst/cycle 200, Peak Watt 200, Time 2400 seconds). The samples are centrifuged at 16000 RCF for 5mins at 4°C and supernatant is collected. The 100 µl of Dynabeads™ Protein G for 200ml culture is washed twice with 1 ml of PBS/BSA using magnetic rack. For proteins tagged with 10X-Flag 25 µl of 1mg anti-Flag antibody (M2-antiflag, Sigma) or For Histone H3 25 µl anti-histone H3 antibody and 75 µl of PBS/BSA is added to magnetic beads and incubated for 4 hours in slow rotation at 4°C. The chromatin fraction was incubated (four tubes per sample) with magnetic beads coated with anti-Flag antibody (M2-antiflag, Sigma) overnight at 4°C for Immunoprecipitation (IP). Beads were pooled and end-repair was carried out using 700 µl of T4 DNA master mix by rotating at 37 °C for 40 min. The T4 DNA master mix was removed using magnetic rack and washing 3 times with ChIP wash buffer 3. A-tailing was carried out using 700 µl of Klenow (3′ > 5′exo-) master mix by rotating at 37°C for 50 min. The Klenow (3′ > 5′exo-) master mix is removed using magnetic rack and washing 3 times with ChIP wash buffer 3. Bridge linker was prepared by annealing Linker-F and Linker-R (HPLC purified (250nmole) from IDT (Integrated DNA Technologies))*:* Bridge linker-F: 5'-/5Phos/CGCGATATC/iBIOdT/TATCTGACT -3'. Bridge linker-R: 5'-/5Phos/GTCAGATAAGATATCGCGT -3'. For proximity ligation, 1.4ml of T4 DNA ligase master mix containing bridge linker is added to the samples and incubated at 16°C overnight. 200 µl of Elution Buffer is added to each tube and incubated at 65°C for 20mins in a shaker. The tubes are centrifuged for 1min at 16000 g at room temperature and samples were

collected using the magnetic rack. Samples are treated with 95 µl of TE SDS, 10 µl of Proteinase K and 6 µl of RNaseA. Reverse crosslinking was carried out at 65 °C overnight. The samples are purified with DNA clean and concentration kit from Zymoresearch and eluted with 15 µl of elution buffer from the kit. Qubit™ 1X dsDNA HS Assay Kit. Tagmentation of proximity ligated DNA was carried by Tn5 transposome by using Nextera® DNA Sample Preparation Kit where 50ng of DNA in 17 µl and 33 µl of tagmentation mix were incubated at 55 °C for 5 mins on thermocycler. The samples are purified with DNA clean and concentration kit from Zymoresearch and eluted with 15 µl of elution buffer from the kit. Purified DNA fragments containing linker DNA were enriched with 25 µl of Dynabeads M-280 Streptavidin. Before adding samples to beads, the beads are washed with 150 µl of ChIP Wash Buffer 5. Beads were suspended with 100 µl of iBlock buffer, incubated for 45 mins in rotation at room temperature and washed once with 200 µl of ChIP Wash Buffer 4. Beads were suspended with 100 µl of ChIP Wash Buffer 4 and incubated for 30 mins in rotation at room temperature and the buffer was removed. 500ng of DNA and 50 µl ChIP Wash Buffer 5 was added to the beads and incubated for 45 mins in rotation at room temperature. The supernatant was removed, DNA in the beads were washed with 500 µl of SSC/0.5% for five times and washed twice with 500 µl of ChIP Wash Buffer 4 and the beads were resuspended with 30 µl of ddH_2O. The sequencing adapters are ligated and PCR is performed for 13 cycles according to standard Illumina sequencing amplification protocol. The PCR products are purified using AMPure XP Bead and Paired end sequencing was performed. ChIA-PET data contains approximately 25 million paired end reads.

3.12 ChIA-PET data analysis

The analysis pipeline is adopted from (Li, Luo et al. 2017) and modified for yeast genome. The paired end raw reads were filtered based on the quality value (-q 20 and –p 30) using FASTX Toolkit. The filtered paired end tag reads were scanned for bridge linker (ACGCGATATCTTATCTGACT, AGTCAGATAAGATATCGCGT) allowing up to two

nucleotide mismatches in the bridge linker using cutadapt. Fastq files were generated with the sequences flanking bridge linker sequences and the linker sequences were removed from the matching reads. The reads were aligned to the reference genome (SacCer 2011) using bwa mem module. Picard Markduplicates module is used to remove the PCR duplicates. The aligned bam file was converted to bed pair end interaction file (bedpe) for cluster generation using bedtools (bamtobed) module. For PET clustering, only the Paired end tags (PET) with distance greater than 1 kb were considered for further analysis. Individual PET interactions were clustered by extending each PET by 500 bp and PETs overlapping at both ends were clustered together as a single PET cluster. PET clusters with the score equal to or more than 2 were considered for meta-analysis.

3.13 Chromatin Conformation Capture Hi-C

The cells were cultured to reach the concentration of 1×10^7 cells/ml in 100ml of culture (50ml culture per sample). Cells were crosslinked with 1% formaldehyde in culture medium for 30 min at room temperature in shaking conditions followed by quenching with 0.125 M glycine for 5 mins in shaking and transferred to ice for 20 mins. Cells were washed with 1ml of Hi-C Lysis Buffer and 400 µl of Hi-C Lysis Buffer was added to the cell pellet followed by glass beads. Cells were lysed with PowerLyzer 24 Homogenizer at 4000 rpm for 20 sec for 8 rounds. In between each round the cells were kept in ice for 5 mins. The samples were collected from glass beads and centrifuged to collect the pellet at 13400 RCF for 10 mins at 4°C. The cell pellet was washed with 1ml 1X TBS buffer and the pellet was dissolved in 1X TBS buffer with 0.1% SDS and incubated at 4 degree in rotation for 30 mins. The samples were incubated at 65°C for 15mins and transferred to ice for 15 mins. 1% Triton was added to samples and incubated at room temperature for 10mins. Samples were centrifuged at 20000g for 10mins at 4°C and the supernatant was discarded. The samples were washed with 500 µl Digestion buffer and dissolved in 200 µl of Digestion buffer and 20 µl of DpnII enzyme was added to samples and incubated at 37°C for 5 hours

with shaking ON for 10seconds, OFF for 1minute 50seconds. Additional 5 µl of DpnII enzyme was added to samples after 2 hours of incubation at 37°C. End filling of digested chromatin was performed with nucleotide dA, dT, dG and biotinylated dCTP. For 250 µl of digested chromatin, 9 µl of each nucleotide with 1mM dA, dT, dG, 20 µl of biotinylated dCTP at 0.4mM, 7.5 µl of Klenow fragment were added to the samples and incubated at 37°C for 1 hour. The samples were stored at 4°C overnight. The enzymes were inactivated by incubating at 70°C for 10 minutes and transferred to ice immediately for 15 minutes. Ligation was performed in 4ml volume with following mixture 250 µl end-filled chromatin, 3340 µl ddH2O, 400 µl T4 DNA Ligase buffer and 25 µl T4 DNA Ligase enzyme were incubated in room temperature for 6 hours. Reverse crosslink was carried out using 10 µl of ProteinaseK and incubated at 65°C for overnight. Purification by Phenol chloroform and ethanol precipitation was performed, 1 volume of Phenol:Chloroform:Isoamyl was added and vortexed for 30sec and allowed to stand at room temperature for 5minutes. MaXtract high density protocol was performed to separate the DNA from proteins and other contaminations. Ethanol precipitation was performed by adding 2 volumes (8 ml) of 100% EtOH and mixed. It was centrifuged for 10 min at 13000g at 4°C and supernatant was discarded. The pellet was rinsed with 5 mL of 70% EtOH. It was further centrifuged for 10 min at 13000g at 4°C and supernatant was discarded. The pellet was air dried and resuspended in 300 µl 1XTE buffer, 6 µl of RNAse and incubated at 37 degrees for 1 hour. Samples were purified with Qiaquick PCR purification kit 3 tubes per sample and eluted with 100 µl of elution buffer for each tube from the kit. The DNA quantified using the nano drop approximately contained 300-400 ng/µl. The samples were treated to remove the biotin in the unligated ends. For the 300 µl of DNA, 66uL NEBuffer 2, 6.6uL BSA (NEB), 6uL 10mM dATP, 6uL 10mM dGTP, 7.5uL T4 DNA Polymerase (NEB) were added and incubated at room temp for 10minutes, then for 1 hour in the PCR machine at 12°C. The samples were purified with DNA clean and concentration kit (2 tubes per sample) from Zymoresearch and eluted with 69 µl of elution buffer for each tube from the kit. The samples

were sonicated using Covaris to reach average fragment length of 300bp. MicroTube AFA fiber 130 µl was used for sonication with the following parameters 17.5 average power, 10 % duty factor, 200 cycles/burst, 180 seconds time, 130 µl sample volume, 6°C Temperature, 6 water level and with intensifier ON. After shearing the samples were purified using DNA clean and concentration kit (1 tube per sample) from Zymoresearch and eluted with 50 µl of elution buffer from the kit. Biotinylated sample and bead immobilization using 25 µl Dynabeads M-280 Streptavidin was carried out. The beads were washed twice with 100 µl ChIP wash buffer 5 and resuspended with 50 µl of ChIP wash buffer 5. 50uL of sheared and purified DNA was added to 50uL of beads. Beads were incubated at room temperature for 1 hour in slow rotation. Short centrifugation was performed and supernatant is discarded using magnetic rack. Beads were washed three times with 200 µl of ChIP wash buffer 5. End repair and dA-Tailinig was performed using NEBNext Ultra II End repair/dA-Tailinig kit. 50 µl of ddH_2O was added to beads followed by 3 µl NEBNext Ultra II End Prep Enzyme Mix and 7 µl NEBNext Ultra II End Prep Rxn Buffer. Samples were gently suspended and incubated in thermocycler for 30 minutes at 20°C, 30 minutes at 65°C and held at 4°C. Illumina index adaptors with barcodes were ligated to the sample. For total volume of 60 µl sample, 6 µl of T4 DNA Ligase Buffer, 1 µl of Illumina barcode adapters and 3 µl of T4 DNA Ligase was added and incubated at room temperature for 10 minutes. Beads were washed three times with 200 µl of ChIP wash buffer 5 and PCR was performed for 13 cycles according to standard Illumina sequencing amplification protocol. The supernatant was collected using magnetic rack, followed by purification and size selection using AMPure XP Beads and paired end sequencing using NextSeq 500/550 High Output Kit v2.5. Hi-C data contains approximately 100 million paired end reads.

3.14 Hi-C data analysis

The paired end raw reads were filtered based on the quality value (-q 20 and –p 30) using FASTX Toolkit. HiCUP toolkit was used for the Hi-C data analysis. The yeast genome

(Saccer2011) was digested with Dpn2 (^GATC) four base sequence using “hicup_digestor” tool to obtain the position of the genome where the Dpn2 cuts. The paired end raw reads were used to find the Hi-C junctions using “hicup_truncator”. The truncated sequences were aligned to the reference genome and then paired using the “hicup_mapper” to produce BAM file. The digested genome file with the Dpn2 restriction enzyme and the aligned BAM file were compared using “hicup_filter” to produce BAM file with true position of Dpn2 digestion. The read duplicates were removed using “hicup_deduplicator”. The final BAM file was used to produce the juicer format (pre) file using “hicup2juicer” conversion script. The “juicer pre” tool was used to produce the .hic file with multiple resolutions (1kb, 5kb, 10kb, 25kb) contact frequency. The “juicebox” was used to visualize the hic map and differential hic maps and custom-made R and python scripts were used for downstream analysis.

3.15 Statistics and Reproducibility

All the experiments were carried out with two biological replicates. To test the significance of the overlap between two replicates (supercoiling, protein and hybrid peak calls), intersect and Fisher exact test from bedtools were used. For bedtools intersect, minimum of 80% overlap were expected for further downstream analysis like meta-gene plotting. The number of overlap peaks and sum of overlap bases between two sets of intervals from bedtools were visualized using VennDiagram library from R. Protein coding genes (n=6706) from saccer2011 were used for meta-gene plotting.

3.16 Code availability

All the R and python scripts used in this work are available in the git repository https://github.com/adhilmd/TopologyCustomAnalysis

4. RESULTS

4.1 DNA topology of gene boundaries and other functional elements

4.1.1 Mapping of DNA supercoil across the yeast genome

Despite being an inherent property of chromatin, only limited studies on DNA supercoil have been performed at whole genome level to understand the mechanistic role of DNA supercoil in chromatin and genome organization. To map the DNA supercoil across the whole genome, Psoralen derivative (4,5′,8-Trimethylpsoralen, TMP) is used to monitor the negative supercoiling, as intercalation of psoralen molecule to DNA is directly proportional to negative super helical tension (Sinden, Carlson et al. 1980) (Figure 1A). We have adopted a previously described method where a biotin molecule was attached to Trimethylpsoralen via a linker (bTMP) (Naughton, Avlonitis et al. 2013) to budding yeast.

To understand the DNA topological distribution across the yeast genome in S Phase of the cell cycle, the cells were arrested using alpha factor and released into fresh media for 15 minutes. Finally, sodium azide (0.1%) was added to metabolically arrest the cells through cytochrome C oxidase inhibition in mitochondria and decrease ATP production (Harvey, Hardy et al. 1999) thereby inhibiting cell division (Ciesla, Mardarowicz et al. 1974). Blocking the cells with sodium azide will help preserve the topological structure by the reduction of genomic transactions such as transcription and replication. bTMP is incorporated along with Dimethyl Sulfoxide (DMSO) to enhance the cell membrane permeability, where bTMP intercalated into the negatively supercoiled regions. This is followed by UV crosslinking (365 nm light) which formed a covalent bond between psoralen and DNA. The covalent bond formed thymidine adducts between the 5,6 double bond pyrimidine bases and 4’5’ furan double bond of psoralen (Kanne, Straub et al. 1982). UV crosslinking was followed by cell lysis, chromatin fragmentation, DNA purification and

biotin enrichment using streptavidin magnetic beads. Finally, the samples were hybridized in whole genome microarray chip.

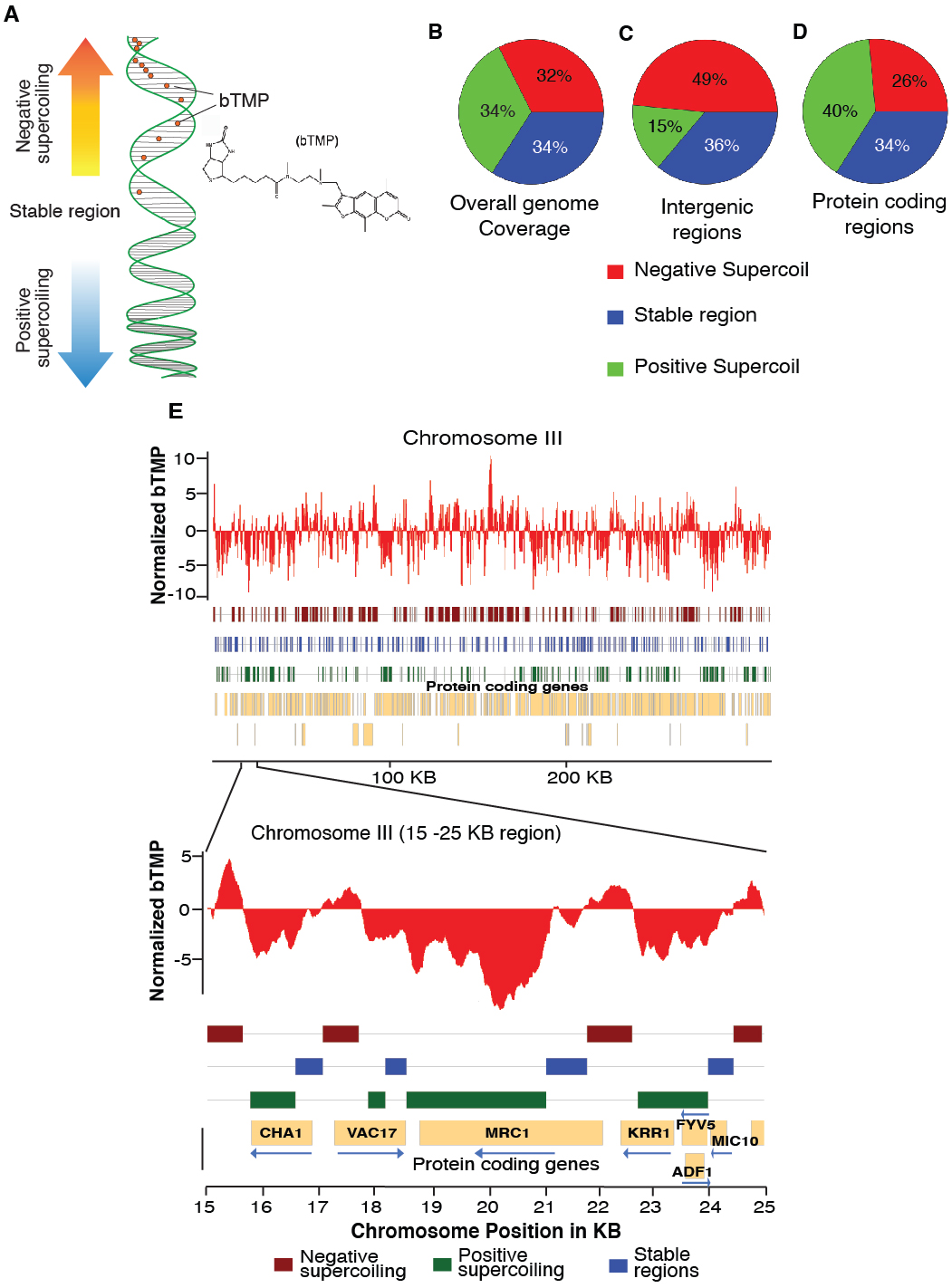


Figure 1: Supercoil base coverage across yeast genome and Chromosome III supercoil distribution A) Schematic representaion of preferential bTMP binding in loosely coiled DNA strands and structure of bTMP probe that interclates into underwound DNA. B) Overall genome supercoil base coverage. C) Intergenic regions (Non transcribing regions) supercoil base coverage D) Protein coding regions (ORF) supercoil base coverage. E) Chromosome 3 view of bTMP (IP/Input) binding score along with three classes of supercoil (Negative, Positive and Stable regions) and enlarged view of chromosome 3 from 15-25kB with supercoil and protein coding genes annotation.

The bTMP ratio was calculated using IP (Biotin enriched DNA) and Input signals (Whole genomic DNA), where higher bTMP ratio represents higher negative supercoil tension. To eliminate the nonspecific binding of bTMP drug, bTMP experiment was performed on naked genomic DNA. The bTMP ratio (IP/Input) from naked genomic DNA was subtracted from the whole cell experiment to obtain normalized signal to noise value.

We categorized normalized bTMP binding ratio into 3 categories i.e., Positive supercoil (over wound), Negative supercoil (under wound) and Stable Regions. Peaks with high bTMP binding ratio (ratio above +1.5) were considered as negatively supercoiled (underwound-DNA) and peaks with low bTMP binding ratio (ratio below -1.5) were considered as positively supercoiled (over wound-DNA). Regions in between cut off values (i.e., -1.5 to +1.5) were considered as stable regions. Chromosome 3 with the three classes of supercoil and enlarged 15-25KB is shown in Figure 1B, along with the pol2 coding genes. The nucleotide base distribution based on these three categories are shown in Figure 1C. The genome is approximately divided into three equal categories based on the base proportion of supercoiling. The Intergenic regions (Non-coding regions) were analyzed for supercoil distribution (Figure 1D), which revealed that half the intergenic region (49%) was negatively supercoiled, 15% was positively supercoiled and 36% was stable regions respectively. Distribution of supercoil in protein coding regions (ORF) is shown in Figure 1E, which is dominated by positive supercoil (40%) and stable regions (34%). Unlike negative and positive supercoiled regions, stable regions have short peaks and doesn't form a 'topological domain', instead, it acts as a transient border between under or over wound domains. Its distribution is similar across all regions (intergenic and protein coding regions) of the genome.

4.1.2 Functional and physical properties of DNA supercoil

To investigate the distribution of negative supercoil in GC or AT rich regions, the nucleotide composition in yeast was compared with the base composition in negative supercoil enriched regions (Figure 2A). In yeast genome we found no preferential binding of bTMP to GC or AT nucleotide bases. This shows there is no bias of negative supercoil formation in GC or AT rich regions across the genome.

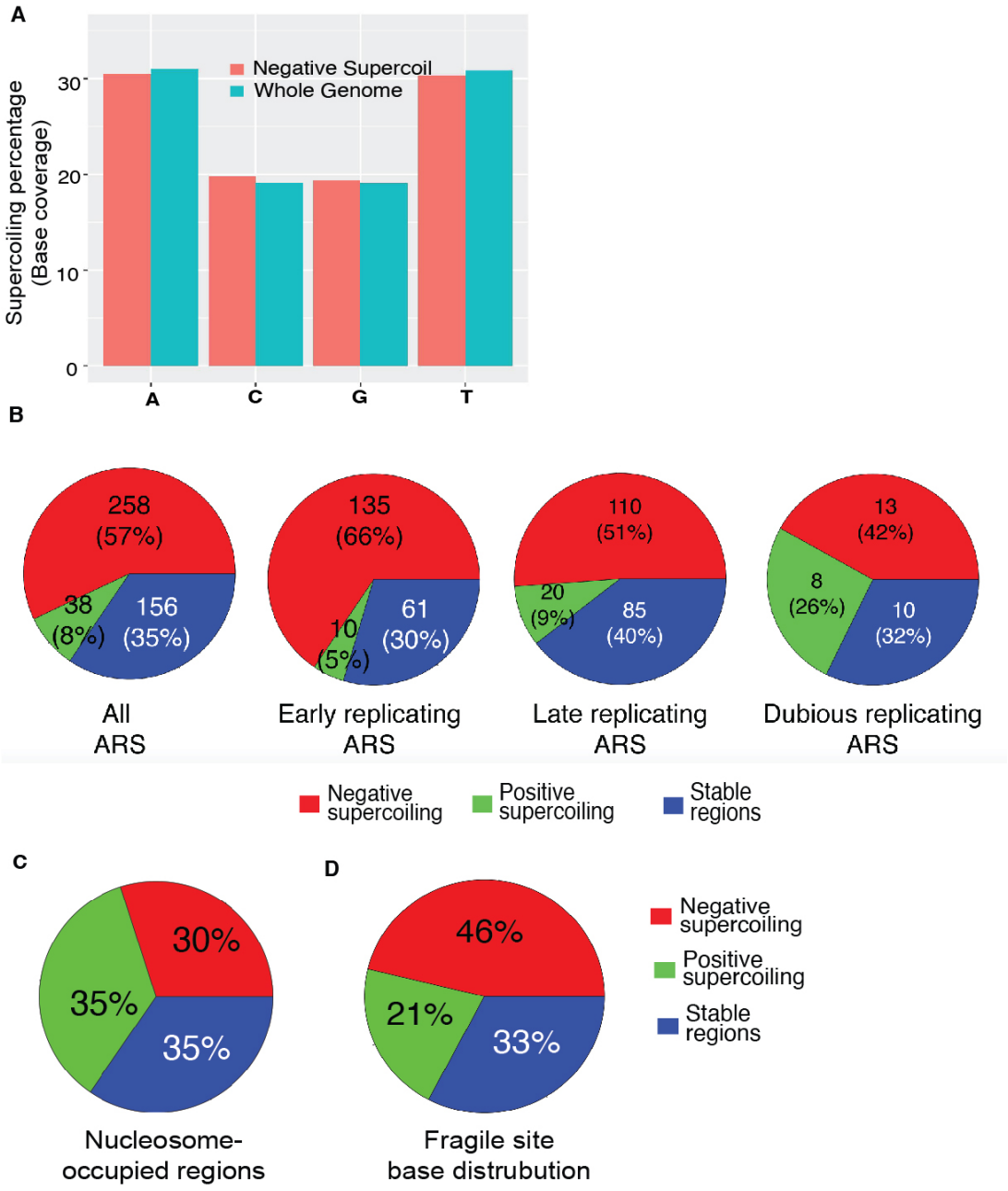


Figure 2: Supercoil base coverage of autonomously replicating sequences (ARS), nucleosome occupied regions and fragile sites. A) Base coverage percentage comparision between negative supercoil regions and whole genome in yeast. B) Percentage of negative, positive and stable supercoil in all, early, late and dubious replication origin sites. Piechart showing the base coverage of negative, positive and stable based on bTMP ratio in C) nucleosome occupied regions and D) Fragile sites across the genome.

In yeast, the chromosomes start to replicate in multiple sites known as replication origin, where each origin initiates two diverging replication forks. The replication origins are specified by DNA sequence motifs and these sequences enable the replication of extrachromosomal plasmids and are thus termed as autonomously replicating sequences (ARS). Interestingly, majority of 'Autonomously replicating sequence' (ARS) accumulate in negatively supercoil regions (Figure 2B). Overall, 258 ARS (57%) are present in negative supercoiled regions, whereas only 38 ARS (8%) are present in positive supercoil regions. Since not all ARS initiate replication simultaneously in eukaryotic cells (Greenfeder and Newlon 1992a, Greenfeder and Newlon 1992b) (Newlon and Theis 1993), we investigated whether supercoil plays a major role between early and late replication. However, we failed to see any such correlation as even late replicating ARS are preferably in the negatively supercoiled regions (110 late ARS - 51%). Interestingly, dubious ARS which are known to initiate only in special cases, has no such preferences for supercoil and are almost equally distributed between the three supercoil regions. From the above observation, replication initiation seems to be influenced by the supercoiled states (Leonard and Grimwade 2010), as negative supercoil facilitate DNA strand opening.

The accumulation of negative supercoil was compared with the nucleosome binding and nucleosome depleted regions. The analysis is performed to study whether the bTMP have any bias on open chromatin where they contain less nucleosome occupancy. The nucleosome binding regions were obtained from published data (Kristin Brogaard 2012) and compared with negative supercoiled regions. In regions containing nucleosome all three negative, positive and stable regions shows equal distributions, proving bTMP binding is independent of nucleosome occupancy (Figure 2C). Highly transcribed regions such as tRNA, LTRs, and rDNA units are considered to be natural fork barriers and are linked to genome rearrangements and chromosome fragility, where we compared supercoil accumulation in fragile sites using previously mapped γ-H2AX enrichment sites (Szilard, Jacques et al. 2010). When we compared base distribution of fragile sites, 46% of the sites tend to have

negative supercoil, whereas 21% and 33% of sites fall in positive and stable regions respectively (Figure 2D). It is still unclear, how DNA supercoil can induce chromosome fragility. One possible reason is requirement of topoisomerase activity to form DNA breaks at fragile sites (Arlt and Glover 2010).

Like many organisms the yeast telomeres are crucial for genome integrity and it consists of nonprotein coding repeats. They also contain sub telomeric, middle and repetitive elements known as telomere associated sequences (TAS) elements. The TAS elements are classified as X and Y' elements, where X elements are heterogeneous in sequence and size and are present in all telomeres. Y' elements are found in zero to four tandem repeats and half of the telomeres lack Y' elements (Dionne and Wellinger 1998) (Klobutcher, Swanton et al. 1981) (Wellinger and Zakian 2012). When the telomeric and sub-telomeric regions were analyzed for negative supercoil enrichment, we found that the regions connecting telomeres and sub-telomeres are highly negatively supercoiled in both short and long telomeres as shown in Figure 3A and 3B.

The DNA sequence of the centromeres are relatively short (± 120 bp) and are essential for proper segregation of the chromosomes by anchoring the chromosomes to the kinetochore during mitosis. Not all the centromeres are enriched with strong negative or positive supercoils, but they are more abundant in stable supercoil regions as shown in Figure 3C. Structurally the centromeres are different in terms of nucleosome occupancy, they are significantly shorter (±125-135bp) than the canonical nucleosome and they are also perfectly positioned in the nucleosome (Cole, Howard et al. 2011). This could be one of the possible reasons for stable supercoil regions around centromeres.

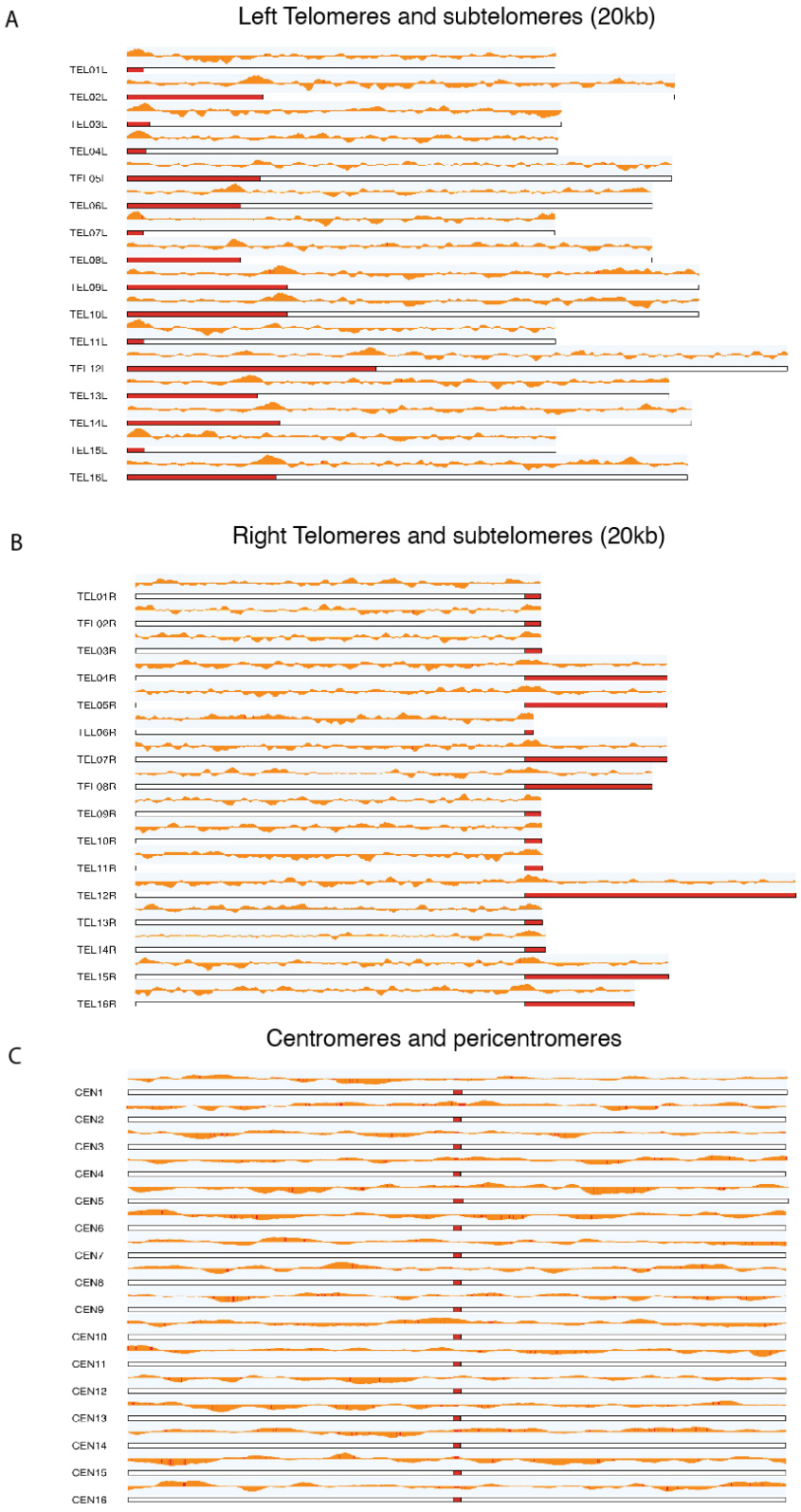


Figure 3: Negative supercoil distribution in telomeres and centromeres. A) Karyoplot showing the negative supercoil distribution in the A) Left telomeres and sub-telomeres (20kb) B) Right telomeres and sub-telomeres (20kb) C) Centromeres and Pericentromeres (5kb).

4.1.3 DNA supercoil state of polymerase I, II, III coding genes

To understand how negative supercoils are distributed in the Polymerase II coding genes we plotted average intensity and gene plot using all the protein coding genes (Figure 4A and 4B). Since median of the number of bases across all the genes in yeast is approximately 1000 bases, we averaged all the protein coding genes to 1000 base pairs along with upstream (500bp) and downstream (500bp) regions. The average intensity is plotted using the normalized IP/Input ratio, whereas for the average gene plot the position is counted when the IP/Input ratio is greater than 1.5. We observed that for both average intensity and average gene count the gene boundaries including the Transcription Start Site (TSS) and Transcription Termination Site (TTS) are negatively supercoiled. Interestingly, we found negative supercoil at TTS and downstream regions. Maximum negative supercoil was observed upstream of TSS (around -250 bp) where promoter opening and assembly of open complex takes place (Nagalakshmi, Wang et al. 2008, Erb and van Nimwegen 2011).

In yeast, approximately 75% of genome is protein coding regions, along with other functional elements like rDNA, tRNA and snoRNAs etc. The protein coding genes are placed close to each other, in fact, in many cases the promoter of one gene overlap with termination region of another gene. To rule out the possibility of bTMP ratio advancing from downstream genes, we selected subset of genes (723 genes) where no neighboring genes were found (500 bp upstream and downstream). Similar to meta-ORF plot of all protein coding genes, meta-ORF plot for these 723 genes (Figure 4C) shows higher amount of negative supercoil in both upstream and downstream regions. Yeast genome also produces a subset of non-coding RNAs including cryptic unannotated transcripts (CUT) and stable unannotated transcript (SUT), often in an antisense orientation (Xu, Wei et al. 2009). We also failed to see a difference in meta-plot, when only 483 protein coding genes, which overlaps with either CUT or SUT were considered, indicating that negative supercoil accumulation downstream or at TTS is an inherent nature of protein coding genes (Figure 4D). Interestingly, these data correspond to a previous study where minichromosomal DNA

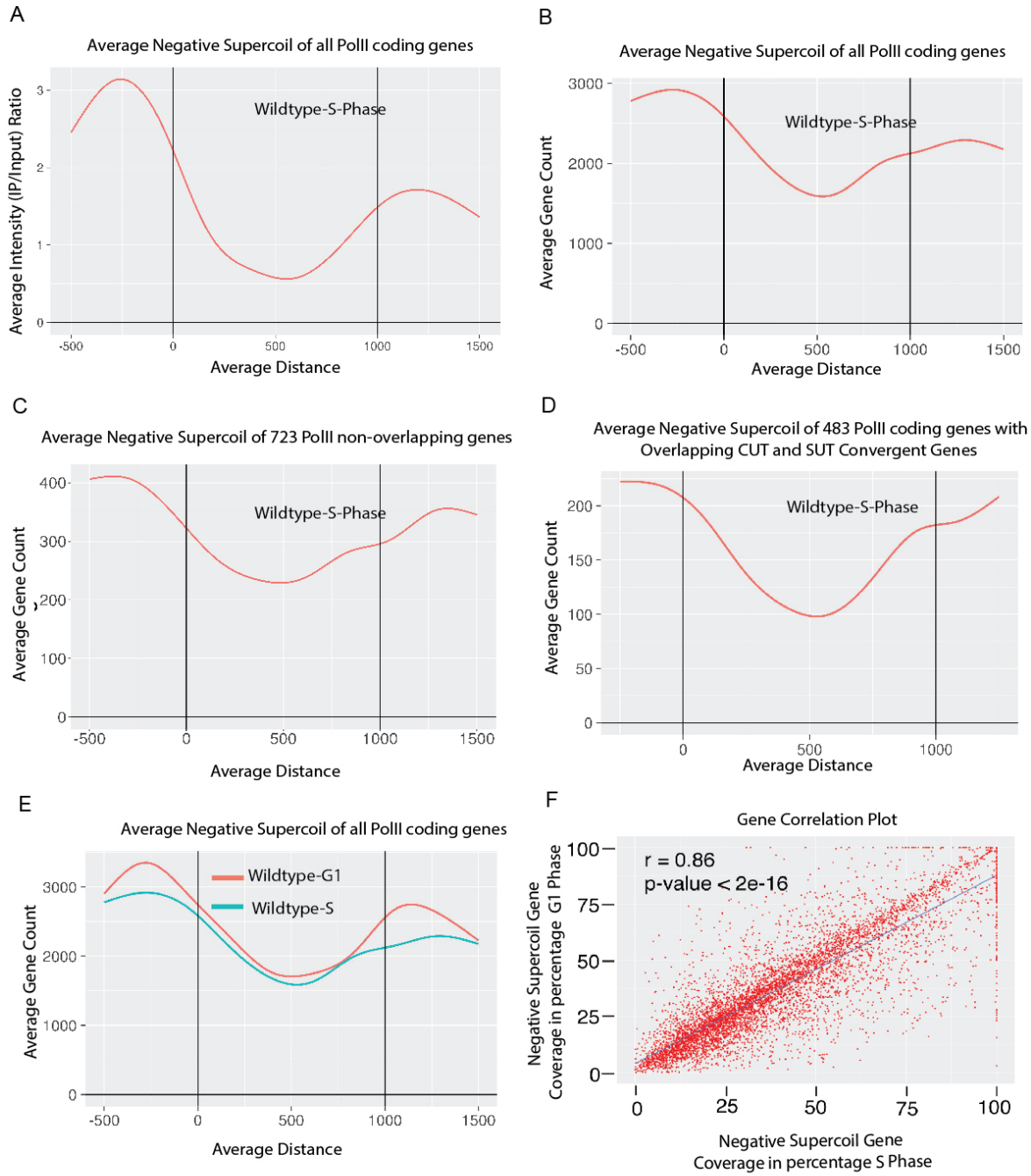


Figure 4: Meta-gene analysis of negative supercoil accumulation in G1 and S phase A) MetaORF intensity (IP/Input) plot of wildtype S phase negative supercoil distribution where all the ORFs are averaged for 1000 bases along with upstream and downstream 500 bases. B) MetaORF gene count plot using the IP/Input ratio threshold >= 1.5 (please refer Materials & Methods section 3.10) of wildtype S phase averaged across all the ORFs C) MetaORF gene count plot where 723 ORFs having no overlapping neighbouring genes D) MetaORF gene count plot where 483 Pol II coding genes with overlapping CUT (cryptic unstable transcripts) and SUT (stable unannotated transcripts) convergent genes E) MetaORF gene count plot of all Polymerase II coding genes between WT-G1 and WT-S phase samples. F) Gene correlation plot between WT-G1 and WT-S phase samples

harboring *REP2* gene showed DNase I sensitivity upstream and downstream but not at the ORF (Lee and Garrard 1991).

We decided to check supercoil accumulation at different cell cycle stages using cells in G1 and S-phase (Figure 4E). For G1-phase, cells were arrested with alpha factor, whereas for S-phase cells were released from arrest by washing out alpha factor and allowing cells

to grow for additional 15 mins. bTMP binding in G1 and S-phase shows similar positive and negative supercoiling accumulation when similar number of peaks were considered. Both negative and positive supercoiling has 89% and 90.2% similarity between cell cycle phases respectively, when common peaks with minimum 10% overlap were considered. When bTMP peaks of G1 and S-phase samples were plotted on an average gene plot, both the curves showed a similar profile (Figure 4E). G1 samples showed a slight increase in negative supercoil, however no such differences were observed with positive supercoil. To analyze the gene correlation, the gene converge of negative supercoil of all polymerase II coding genes (overlapping bases of bTMP peaks with Polymerase II coding regions) for G1 and S was calculated. The gene correlation of bTMP base coverage between G1 and S phase sample was positively correlated and highly significant ($R = 0.86$ and p-value $< 2e\text{-}16$) (Figure 4F).

Along with protein coding genes, other functional genomic regions which produces non-protein RNA's also show very high bTMP peak accumulation. Unlike other RNA producing genes, rDNA which is transcribed by RNA pol I and III showed very interesting topology distribution. The intergenic spacer which contains the replication origin and replication fork barrier are highly negatively supercoiled. As described previously, rRNA genes have high negative supercoiling (Schultz, Brill et al. 1992, French, Sikes et al. 2011) at the intergenic spacer or promoters which is between 35S precursor rRNA and 5S rRNA coding regions, whereas positive supercoiling accumulates at the rDNA transcribed regions. Stable regions which are found at the border between negative and positive supercoiling at different parts of the genome were completely absent at rDNA locus (Figure 5). This sharp peak distribution between negative and positive supercoiling can be attributed to topoisomerase function, as topoisomerase increases elongation efficiency by relaxing positive supercoiling at ORF, and also maintains the pool of negative supercoiling at the promoter to induce RNA pol I transcription (French, Sikes et al. 2011).

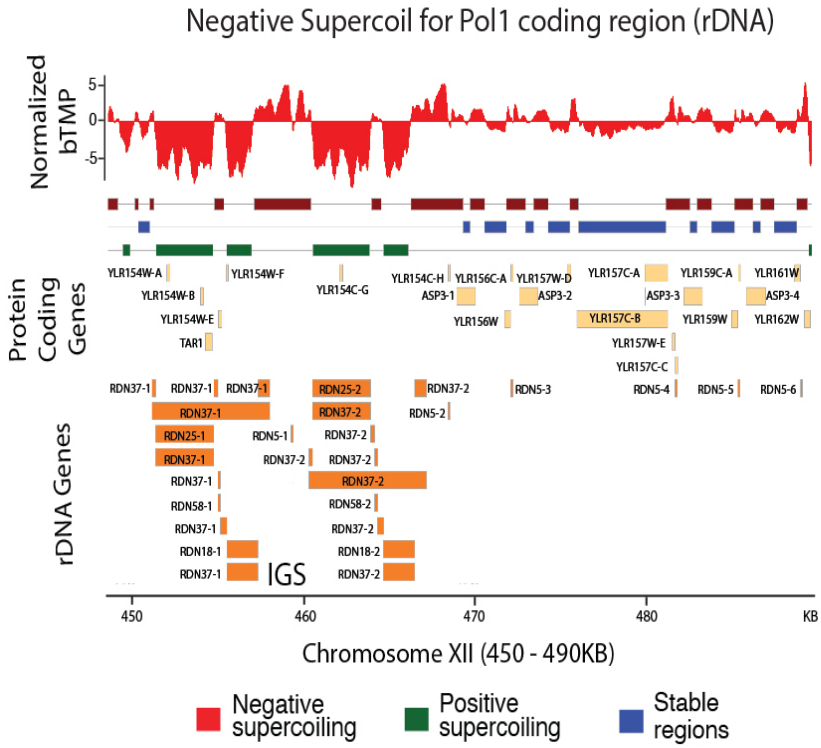


Figure 5: Supercoil distribution in rDNA region (ChrXII 450-490Kb), where rDNA contains two rDNA units with intergenic spacer (IGS) but the normal wildtype contains ~150 copies of rDNA units.

The polymerase III coding regions are significantly negatively supercoiled, particularly tRNA genes. The three tRNA genes (tN(GUU)G, tR(UCU)G3, tI(AAU)G) showing high accumulation of negative supercoil are shown in Figure 6A. The tRNA are small non-coding RNAs that deliver amino acids to the ribosome for protein synthesis. In yeast there are 275 annotated tRNA regions and they are approximately 70 to 100 bases long. In active yeast cells, tRNAs represent approximately 15% of total RNA, indicating the tRNAs are highly transcribed regions in the genome. The average intensity plot of all 275 tRNA genes is shown in Figure 6B. The upstream region of tRNA genes is highly negatively supercoiled compared with the downstream region.

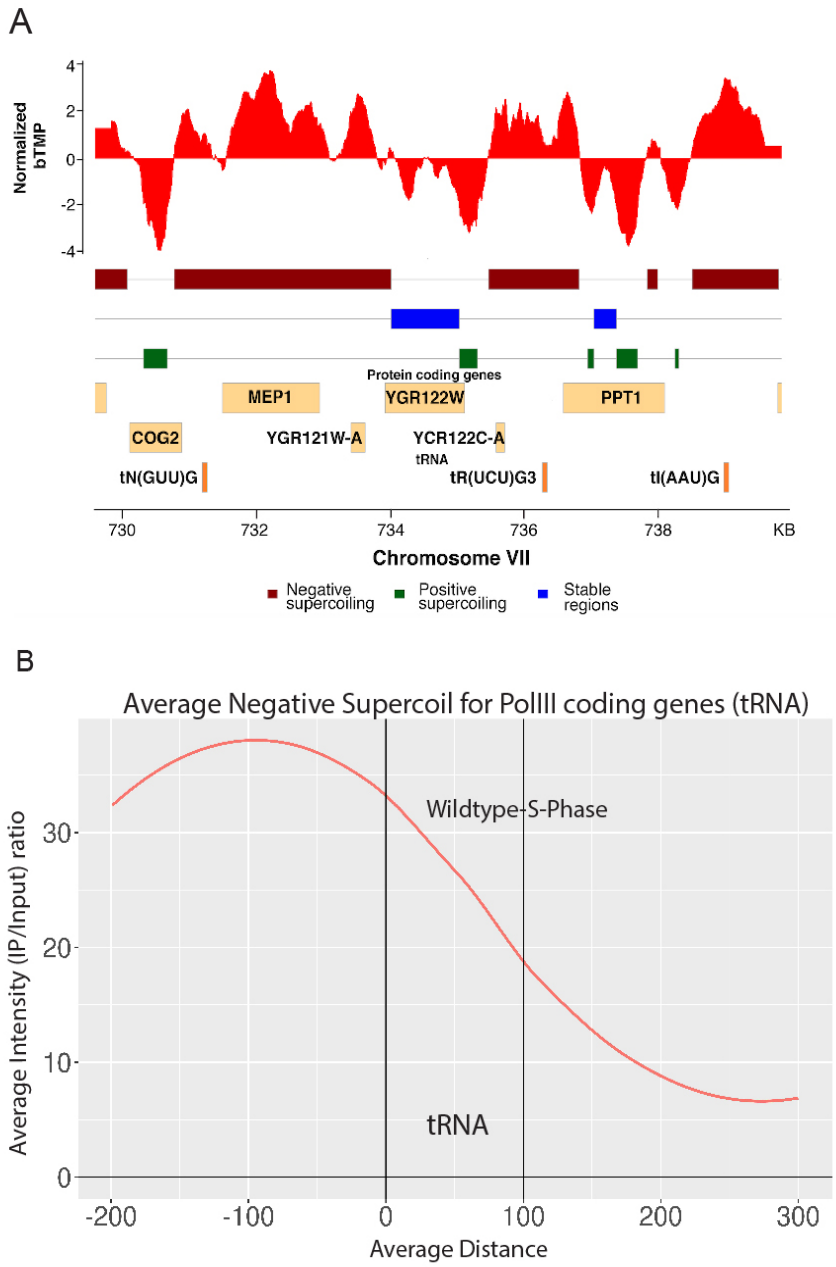


Figure 6: Supercoil distribution in tRNA region A) Chromosome VII region from 730kb to 740KB showing tRNA genes B) MetaORF intensity (IP/Input) plot of negative supercoil distribution for all Polymerase III coding tRNA, where 275 tRNA are averaged for 100 bases along with upstream and downstream 200 bases

4.1.4 The accumulation of negative supercoils correlates positively with gene expression level but not with presence/absence of the gene

In yeast the promoter is a crucial component in controlling gene expression, where most of the regulation takes place at the transcriptional level. The transcriptional level of many

genes can either be increased or decreased by providing a particular carbon source such as glucose or galactose. To see if the supercoil changes under different growth conditions, we compared bTMP chip profile for cells grown in two different carbon sources: glucose and galactose. The cells are grown separately in glucose and galactose respectively and allowed to reach the optimum concentration, after which the cells were blocked with alpha factor and released into S phase for 15 mins. Subset of genes in budding yeast which only expressed in the presence of galactose (GAL genes: GAL1, GAL7 etc) also showed negative supercoil accumulation when grown in glucose or galactose containing media (Figure 7A). Additionally, some genes in yeast are expressed at particular stage of cell cycle, like ASF2 gene which is S-phase specific transcribing gene. We analyzed the negative supercoil distribution from G1 arrested and S phase released cells but failed to find any significant difference in the accumulation of negative supercoil in ASF2 between these two conditions (Figure 7B). This showed that supercoil accumulation was not dependent on transcription status of the individual gene per se.

To see the correlation between gene expression rate and negative supercoil, we carried out RNA-seq from cells in S-phase. After the aligning the read to the yeast reference genome, the aligned reads were counted for each gene. To counter the bias in gene length and differences in sequencing depth, FPKM (fragments per kilobase of transcript per million mapped reads) values were calculated for all the protein coding genes. The genes were categorized based on the FPKM values into three equal categories: high, medium and low expressed genes (Figure 7C). Highly expressed genes accumulate higher negative supercoiling particularly in the upstream and TSS, where assembly of open complex and transcription initiation occurs (Figure 7D). However, no major differences were observed between middle and low expressed gene classes. Surprisingly, the difference in negative supercoil accumulation between highly expressed and the other two classes were only in the upstream, as it is crucial for the formation of transcription initiation complex (Tabuchi and Hirose 1988). Failure to see increase in downstream confers the previous notion, that looping

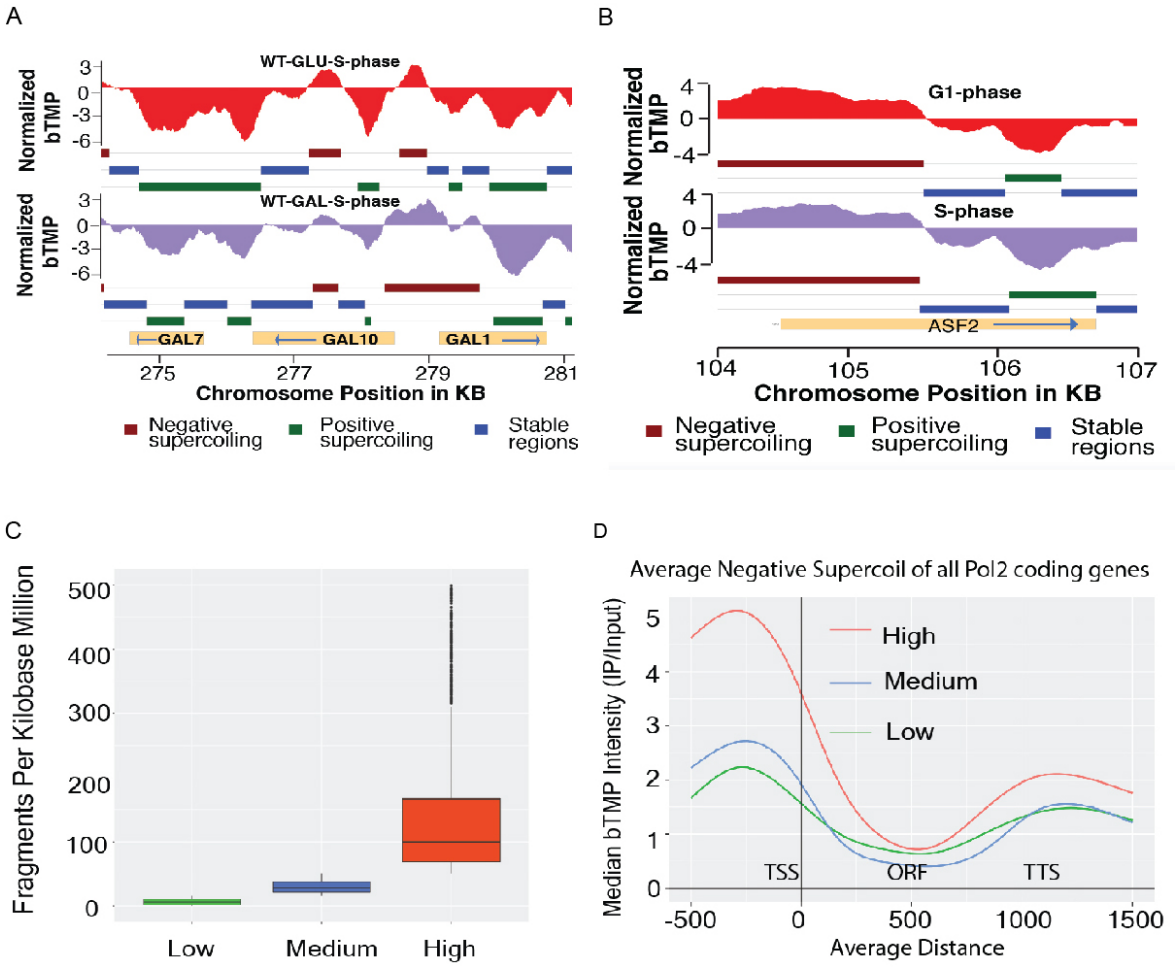


Figure 7: Negative supercoil comparison with transcription. A) Normalize bTMP ratio (IP/Input) showing negative supercoil in Wildtype Glucose S Phase and Wildtype Galactose S Phase in GAL promoter genes. B) Normalize bTMP ratio (IP/Input) showing negative supercoil in Wildtype G1 Phase and Wildtype S Phase in ASF2 gene C) RNA-seq FPKM (Fragments Per Kilobase Million) gene expression categorised into three equal classes (n=6706 genes; Low=Medium=High=2235 genes) D) MetaORF intensity (IP/Input) plot of wildtype S phase negative supercoil distribution for three gene expression classes (Low, Medium and High) from RNA-seq data.

is a general phenomenon of transcription and is not restricted to a particular class or expression levels (Krivega and Dean 2012) (Tan-Wong, Zaugg et al. 2012). Hence, underwound DNA at Pol II gene boundaries were enhanced in highly expressed genes. The highly underwound regions will lead to tertiary structures in DNA, such as cruciform like structures where the same strand tends to coil instead of opposite strand coiling (B-form DNA).

4.1.5 Supercoil architecture based on neighboring gene pair transcription directionality

The yeast genome is densely packed with protein coding genes, where intergenic spaces are only ±25% compared to ±75% of ORF regions. The minimal regulatory spaces poses a great challenge for efficient transcription across the different stages of the cell cycle. In few cases the promoters overlapped or had zero distance with the neighboring genes. The bi-directional promoters have a bi-modal length distribution i.e., short and long bi-directional promoters. Some of the short bidirectional promoter regions are evolutionally conserved across the species (Chen, Wei et al. 2011). Since negative supercoil accumulated more in the intergenic spaces of the genome, we compared the supercoil context at intergenic spaces with respect to gene orientation by grouping Pol II genes into codirectional (+ strand; n=1453 gene pairs and – strand; n=1415 gene pairs), converging (n=1590 gene pairs) and diverging (n=1512 gene pairs) classes. Further the gene pairs were grouped based on the length of the intergenic spaces (<250 bp = 1729 gene pairs, 251-500bp = 2224 gene pairs and >500 bp = 2010 gene pairs). Intergenic spaces between converging genes were smaller compared to the other directional classes (Figure 8A). Divergent genes exhibited larger intergenic spaces. The codirectional and divergent genes have similar negative supercoil and positive supercoil distribution. Converging genes accumulated more positive supercoil at intergenic spaces at the expenses of negative supercoil (Figure 8B and 8C). The intergenic spaces of the convergent genes have only an average of ~326 base pairs while the intergenic spaces of the divergent genes have ~618 base pairs. During transcription, the limited intergenic space in convergent genes allows only ~160 base pairs per gene for proper termination outside of the neighboring ORFs. We propose that convergent gene pairs have specific topological context where it contains more positive supercoil accumulation in the

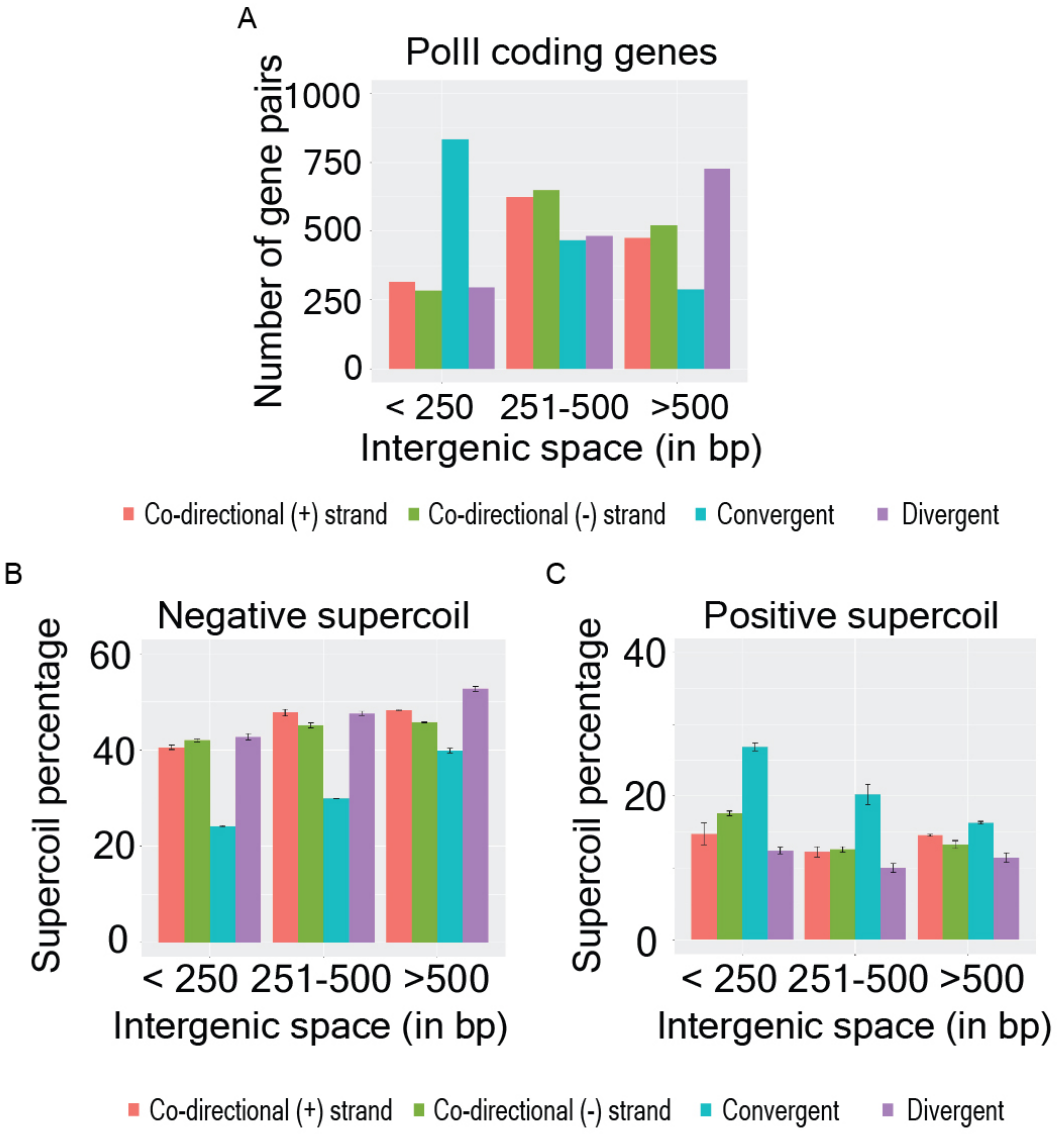


Figure 8: Supercoil distribution based on gene pair transcription directionality A) The Polymerase II coding neighbouring genes are categorized based on the gene pair transcriptional directionality (Co-directional (- strand), Co-directional (+ strand) and Diverging and Convergent) and length of intergenic spaces (<250bp, 251-500bp and > 500bp). B) Base coverage percentage of negative supercoil accumulation with respect to gene pairs grouped according to transcription directionality and different length of intergenic spaces (2 replicates, mean ± s.d) C) Base coverage percentage of positive supercoil accumulation with respect to gene pairs grouped according to transcription directionality and different length of intergenic spaces (2 replicates, mean ± s.d)

intergenic spaces to slow down the polymerase II movement and assist in the polymerase II termination.

4.2 Topoisomerase Top1 and Top2 role in maintaining the DNA supercoil structure

4.2.1 Top1 accumulates along gene bodies, whereas Top2 accumulates at gene boundaries

DNA topoisomerases I and II are involved in DNA relaxation where they catalyze strand passage thus changing the linking number of DNA molecules and resolving the topological problems generated during replication and transcription (Liu and Wang 1987) (Champoux 2001). To study the protein binding profile of Topoisomerases I and II across the yeast genome, the proteins are tagged at C-terminus with either 10X-Flag or 6X-PK respectively. Chromatin immunoprecipitation was carried out using Dynabeads protein G beads coated with anti-Flag antibody or anti-PK antibody. Microarray was carried out using whole fragmented DNA (input) and protein immunoprecipitated DNA (IP) for each sample.

Top2 binds at the intergenic region particularly upstream and downstream of the gene (Bermejo, Capra et al. 2009) (Figure 9A). Top1 protein accumulates inside the ORF, more towards the transcription termination site of the gene (Figure 9B). The number of genes containing Top1 protein binding peaks are less compared to Top2 protein binding peaks. This could be due to colocalization of multiple Top2 protein in short intergenic spaces as compared to protein coding regions. Previous observations indicate redundancy in the functioning of eukaryotic Top2 and Top1 topoisomerases, as both could resolve negative and positive supercoiled DNA (Wang 2002). Moreover, functioning of either one of the topoisomerases is sufficient to ensure the normal progression of DNA replication forks (Kim and Wang 1989, Bermejo, Doksani et al. 2007). Similarly, either Top1 or Top2 could relax positive supercoil generated during elongation of RNAP complex (Mondal and Parvin 2001, Garcia-Rubio and Aguilera 2012, Fernandez, Diaz-Ingelmo et al. 2014, Baranello, Wojtowicz et al. 2016).

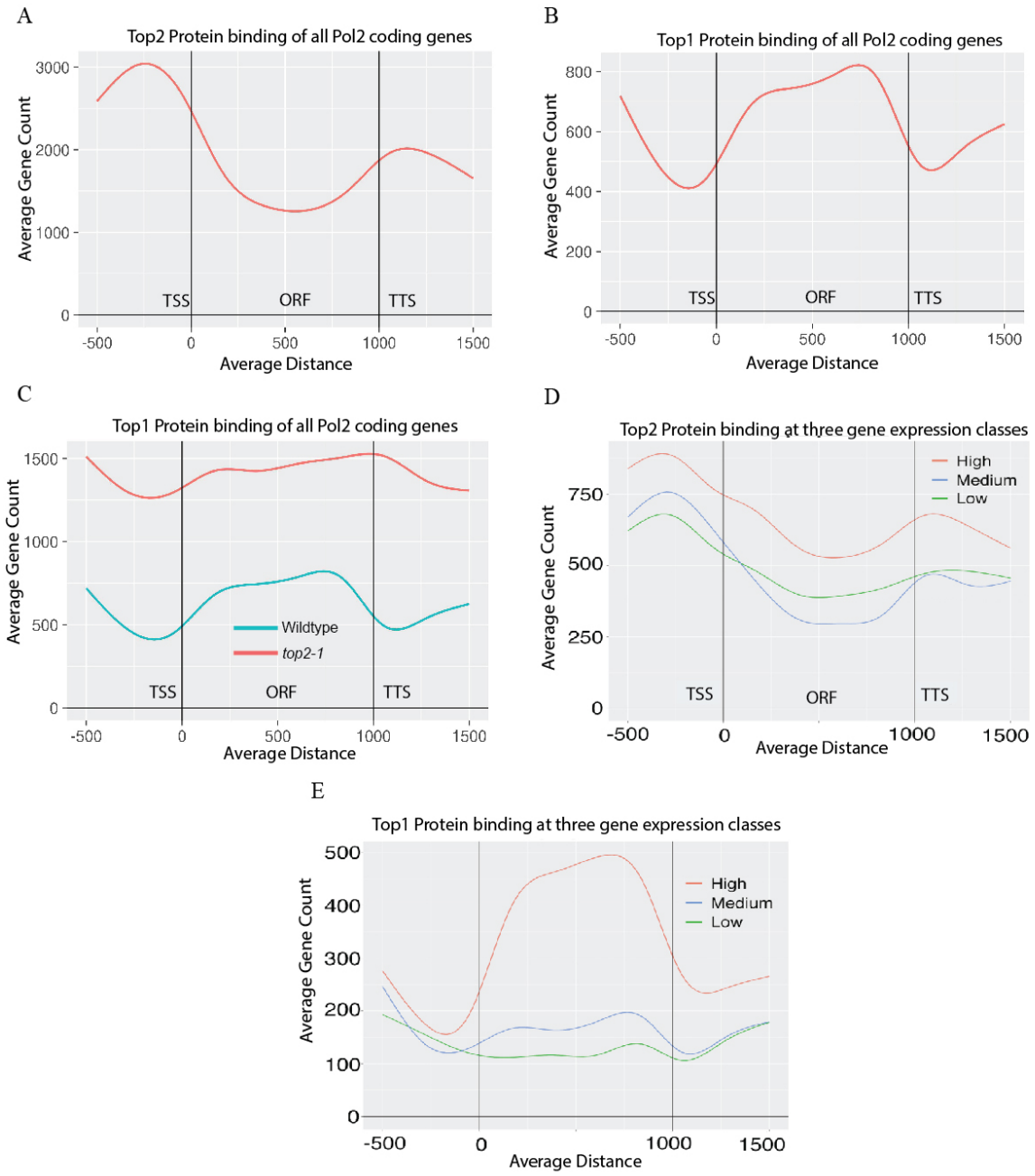


Figure 9: Top1 and Top2 protein binding profile across the genome. MetaORF gene count plot of A) Top2 binding profile averaged across all the ORFs for 1000 bases along with upstream and downstream 500 bases B) Top1 binding profile averaged across all the ORFs for 1000 bases along with upstream and downstream 500 bases C) Top1 binding profile in wildtype and *top2-1* condition where all the ORFs averaged for 1000 bases along with upstream and downstream 500 bases. D) MetaORF gene count (IP/Input) plot of Top2 distribution for three gene expression classes (Low, Medium and High) from RNA-seq data. E) MetaORF gene count (IP/Input) plot of Top1 distribution for three gene expression classes (Low, Medium and High) from RNA-seq data

The activities of Top1 and Top2 are exchangeable (Champoux 2001, Wang 2002). We hypothesized that Top1 could be taking the role of Top2 in *top2-1* cells. The Top1 protein chip was carried out in the *top2-1* background, where the cells are moved to restrictive temperature (37°C) after reaching the required concentration to repress Top2

activity. Top1 protein chip in wildtype background had ~2031 peaks as compared to ~4010 peaks in *top2-1* background. This shows that Top1 protein, increased its accumulation by 2-fold in the absence of Top2 protein (Figure 9C). Interestingly, Top1 increases both in gene boundaries and ORF in the absence of Top2 protein, the binding profile is similar to wildtype. Top1 binding profile in *top2-1* cells does not specifically increase in the gene boundaries to take over the loss of Top2 activities. The Top1 DNA strand rotation mechanism cannot compensate the Top2 DNA cross-inversion mechanism. This is also consistent with previous observation (Joshi, Pina et al. 2012), where Top1 cannot directly compensate for loss of Top2, but can assist transcription machinery by relaxing positive supercoiling. Top2 accumulates in gene boundaries similar to negative supercoil accumulation, whereas Top1 accumulates in ORF like positive supercoil.

We analyzed the similarity between different classes of gene expression rate (Low, Medium and High) from wildtype RNA-seq experiment (Figure 7C) based on the topoisomerases (Top1 and Top2) protein binding profile. The Top2 binding showed less difference between all three classes of gene expression (Figure 9D), where they have similar Top2 protein binding profile in the gene boundaries of all three classes of gene expression. Interestingly, the Top1 enrichment was significantly more with high expression genes (Figure 9E) in the gene boundaries and showed no difference between low and medium expressed genes. Top1 activity is regulated by RNA polymerase II (RNAPII) (Baranello, Wojtowicz et al. 2016) where it has been reported that Top1 pause-release mechanism is stimulated by BRD4-dependent phosphorylation of RNAPII. It is consistent with our observation, where the high accumulation of Top1 in the highly expressed genes results in higher activity of RNAPII.

Another important role of Topoisomerase is maintaining the rDNA stability. It has been shown that in topoisomerase double mutant *(top1Δ top2-4)* half the rDNA copy numbers present as extrachromosomal circles contain one rDNA unit (9kb) or multiple copies (Kim and Wang 1989). rDNA is a highly transcribing region where topoisomerases

Top1 and Top2 relive the supercoil tension generated during the rDNA transcription (French, Sikes et al. 2011). In our data, we found that Top2 binds to rDNA, whereas Top1 accumulation is very low, but in the absence of Top2, Top1 significantly accumulates across the entire rDNA unit (9kb) (Figure 10).

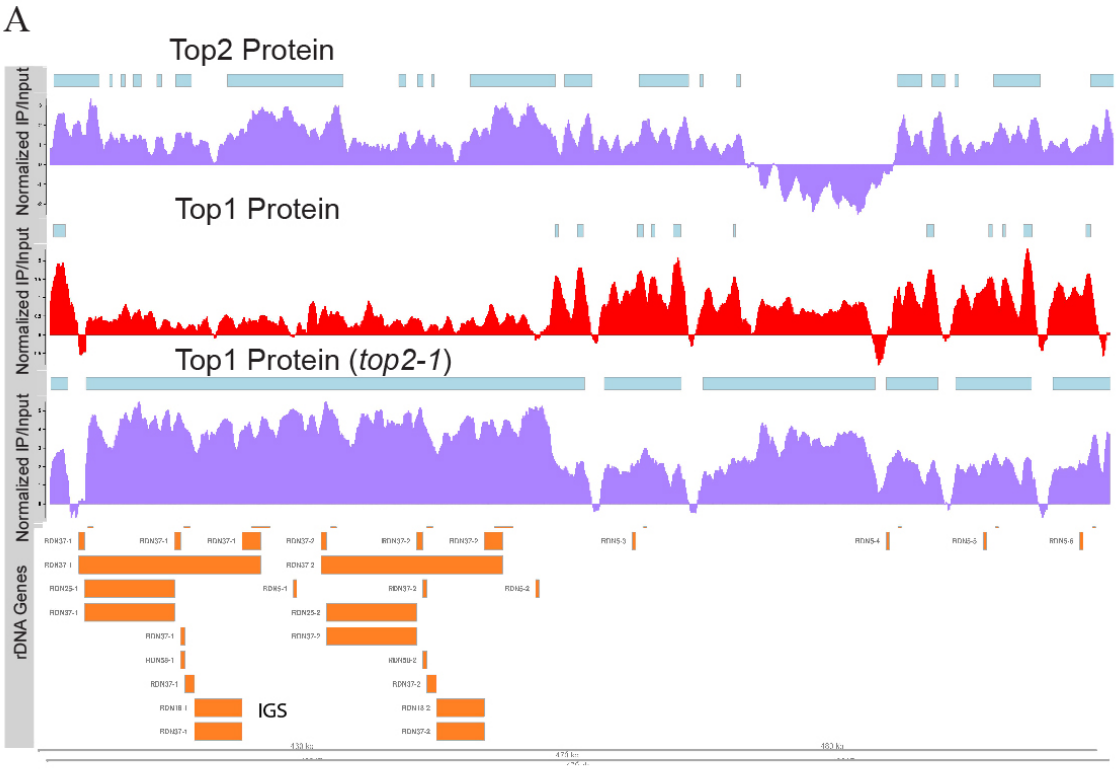


Figure 10: Top1 and Top2 protein distribution in the rDNA region (ChrXII: 450-490KB), where rDNA section contains two rDNA units with intergenic spacer (IGS) but the normal wildtype contains ~150 copies of rDNA units.

4.2.2 Top2 and Top1 topoisomerase's role in maintaining supercoil architecture

To map the negative supercoil using bTMP in *top1Δ,* the cells were cultured at 28^0C, followed by alpha factor arrest to block cells at G1 phase and released into S phase for 20 minutes, whereas for mapping negative supercoil in *top2-1* the cells were cultured in permissive temperature (25^0C) and alpha factor was used to arrest the cells in G1 phase and released into restrictive temperature (37^0C) to enter S phase and the cells were collected after 20 mins. *Top2-1* compared with the wildtype shows 66% common nucleotide base coverage based on the bTMP enriched peaks. To analyze the gene correlation, the gene coverage of

negative supercoil of all polymerase II coding genes (overlapping bases of bTMP peaks with Polymerase II coding regions) for wildtype and *top2-1* was calculated. The gene correlation for bTMP base coverage between wildtype and *top2-1* sample have a low correlation value (R=0.70) (Figure 11A). The *top1Δ* cells compared with wildtype shows higher common nucleotide base coverage (84.4%) and high gene correlation value (R=0.87) based on the bTMP enriched peaks (Figure 11B).

Meta gene plot for *top2-1* cells shows significant reduction in negative supercoil accumulation in both upstream and downstream regions (Figure 11C). In upstream and downstream regions, average peak value reduced significantly compared to the wild type, however no such alterations were observed within the ORF region. No significant changes were observed in *top1Δ* cells (Figure 11D), probably explaining why *top1Δ* mutants are viable in yeast having no effect on replication fork progression and showing minor differences in gene expression (Bermejo, Doksani et al. 2007).

Top2 mainly localizes upstream and downstream of a gene and maintains a pool of negative supercoil by avoiding accumulation of positive supercoil. On the other hand, Top1 relaxes positive supercoil by associating with RNAP complex in the ORF during elongation. To see the effect of reduction in negative supercoil, we selected genes showing more than 40% difference in bTMP binding in *top2-1* cells compared to wild type. Functional enrichment analysis with genes showing difference either in upstream (785 genes) or downstream (649 genes) showed no specific biological processes or pathways. Previous reports where *top2* was inactivated for 120 mins, suggests Top2 is crucial for transcription of longer protein coding genes (Joshi, Pina et al. 2012), particularly genes which are above 3 kB. However, we failed to see any dependency on gene length and reduction in negative supercoil in *top2-1*cells. This difference in two independent studies could be due to the difference in duration of *top2* inactivation as in this study we used 20 minutes rather than 120 minutes as the time for Top2 inactivation.

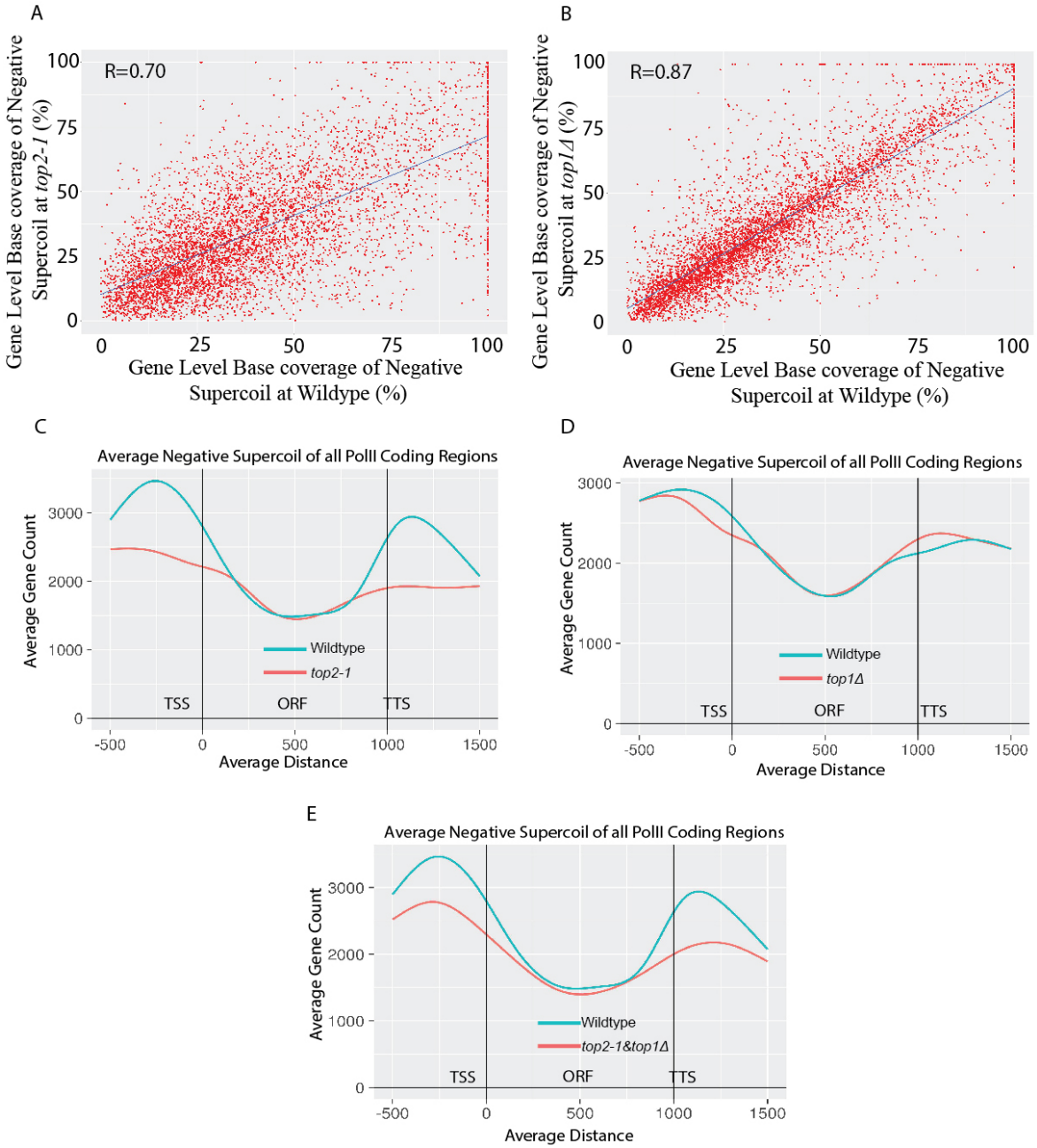


Figure 11: Negative supercoil distribution in the absence of topoisomerases Top1 and Top2 A) Gene correlation plot of negative supercoil between wildtype and *top2-1* B) Gene correlation plot of negative supercoil between wildtype and *top1Δ* C) MetaORF gene count plot of negative supercoil in wildtype and *top2-1* averaged across all the ORFs for 1000 bases along with upstream and downstream 500 bases D) MetaORF gene count plot of negative supercoil in wildtype and *top1Δ* averaged across all the ORFs for 1000 bases along with upstream and downstream 500 bases E) MetaORF gene count plot of negative supercoil in wildtype and *top2-1&top1Δ* averaged across all the ORFs for 1000 bases along with upstream and downstream 500 bases.

Top2 and Top1 work in coordination to produce RNAP2 transcript and global RNA synthesis reduction was only observed in *top1Δ top2-1* double mutants in yeast (Gartenberg and Wang 1992). To test the supercoiling accumulation in topoisomerase double mutant, we carried out bTMP-chip in *top2-1&top1Δ*. Since double mutants failed to enter S-phase at restrictive temperature (Bermejo, Doksani et al. 2007), cells were released to S-phase at

permissive temperature (25^0C) and later moved to restrictive temperature (37^0C) to inactivate Top2. In double mutant, negative supercoil accumulation was reduced when compared to wild type (Figure 11E).

4.2.3 Top2 mediate chromatin loop formation

Previously we had hypothesized that Top2 and Hmo1 might play a major role in chromatin loop formation in yeast (Bermejo, Capra et al. 2009). To identify the protein that is directly involved in chromatin-chromatin interactions, we adapted long-read ChIA-PET (Chromatin interaction analysis by paired-end tag sequencing) method to yeast (Li, Luo et al. 2017). Cells where, Top2 and Hmo1 genes that were tagged at C-terminus with 10X-Flag or 6X-PK respectively were used for ChIA-PET experiments. After sequencing, 7.3 and 8.8 million paired end reads were obtained, which gave 0.3 and 0.9 million uniquely mapped paired end tag sequencing (PETs) for Hmo1 and Top2 respectively. This resulted in two genome-wide datasets: ChIP enriched protein binding site and protein bound chromatin interactions. To eliminate the self-annealed PETs during the ligation step of ChIA-PET procedure, we fixed a minimum distance of 1 kilobase for protein bound chromatin interactions (Figure 12A). Interestingly, only Top2 samples showed considerable amount of inter ligation PET clusters (2315). Among which, 1983 (85%) were intra-chromosomal interactions whereas 332 (15%) were inter-chromosomal interactions. Although 70% of Hmo1 chromatin-binding site matches with Top2 binding sites (Bermejo, Capra et al. 2009), only 75 chromosomal interactions were obtained from Hmo1 ChIA-PET. Since Hmo1 samples showed lower number of interactions we speculate that Hmo1 may not be directly involved in chromatin-chromatin interactions but might assist in Top2 mediated chromatin interactions. Unlike higher eukaryotes, inter-chromatin loops in yeast formed by Top2 proteins were smaller in size, in the range of 1.5kb to 2.5kb average size being 1800 bp (Figure 12B). Although loop length varies in size, majority of the loops fall between 1500 to 2000 bp size.

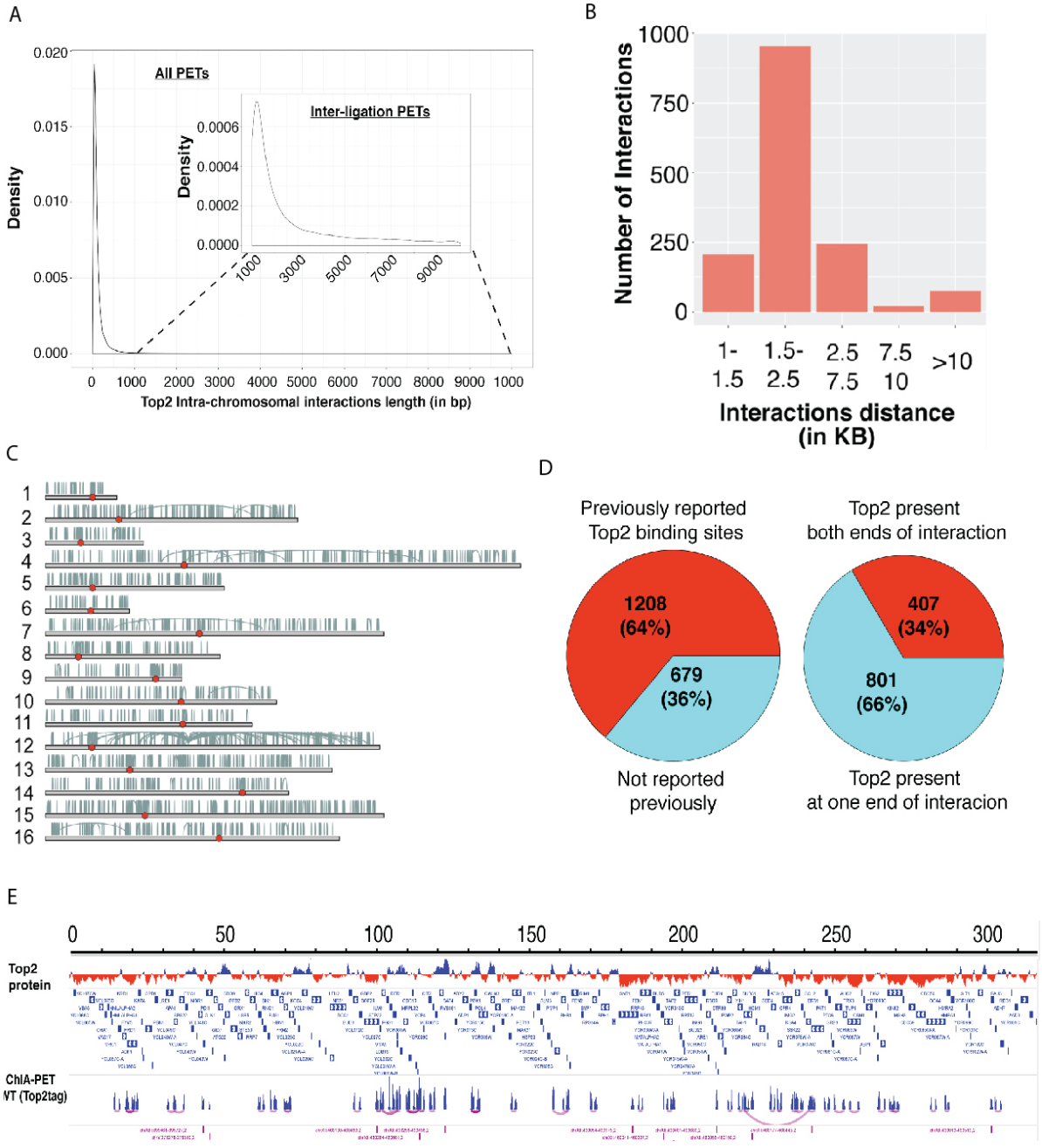


Figure 12: Chromatin loops mediated by Top2 protein in yeast A) Density plot showing all paired end tag sequencing with the length of interaction from ChIA-PET B) Bar plot showing the number of interactions in each interval (1-1.5kb, 1.5-2.5kb, 2.5-7.5kb, >10kb) C) Karyoplot showing the interaction clusters in each chromosome (red circle indicating the centromere in each chromosome) D) Comparision of Top2 protein binding with Top2 interaction from ChIA-PET E) All interaction in chromosome 3 from ChIA-PET with Top2 binding regions.

There are several multiple clusters of intrachromosomal interactions, among which centromeres have many such clusters of interactions, whereas telomeres have very few interactions (Figure 12C).

To validate the chromatin interactions obtained from Top2 samples we compared ChIP- enriched binding site dataset obtained from ChIA-PET with previously mapped protein ChIP-CHIP data (Bermejo, Capra et al. 2009). 1583 of 2315 (68%) chromatin interactions coincide with previously reported Top2 binding sites, whereas remaining 732 chromatin interactions were mapped to previously unknown Top2 binding sites (Fig 12D). Only 38% chromatin interactions contain previously reported Top2 binding site on both the ends of interaction whereas rest 62% of interactions contains Top2 binding site only at one end. Furthermore, chromatin loops also exist as a cluster, where several smaller chromatin loops are placed within another loop (Figure 12E). Although it is not clear why only some loops exist as a cluster and others as a simple loop, it is of interest to note that intra-chromosomal interactions often exist in these clusters. Altogether, our ChIA-PET data demonstrate that Top2 protein mediates both inter and intra-chromosomal interactions during S-phase, which might act as a topological barrier to avoid the replication and transcription conflict.

4.2.4 Ablation of Negative Supercoil by over expressing E.coli DNA topoisomerase I (TopA) in yeast topoisomerase double mutant

We showed that wild type cells accumulate negative supercoiling at flanking regions of ORFs. To further validate the genomic observations based on the bTMP technique we expressed *E.coli* DNA topoisomerase I (TopA) in yeast cells. It has been shown that the simultaneous inactivation of yeast topoisomerases I and II, combined with the expression of *E.coli* DNA topoisomerase I (TopA) removes negative supercoil in plasmids and induces positive supercoiling (Gartenberg and Wang 1992, Trigueros and Roca 2002). Moreover, significant reduction of mRNA level was observed in topoisomerases double mutant when E.coli TopA was expressed. Wild type and topoisomerase *top1Δ top2-1* double mutants containing either empty vector (control plasmid) or *E.coli* TopA expressing plasmids were grown at permissive temperature and moved to restrictive temperature for 120 mins and

supercoil accumulation was measured by bTMP-chip protocol. The cells were unsynchronized and FACS profile revealed that, wild type cells had a heterogeneous cell cycle distribution in both permissive and restrictive temperature, whereas the topoisomerase mutants with TopA expression had a heterogeneous cell cycle distribution in permissive temperature but were blocked in G1 phase under restrictive temperature (Figure 13A).

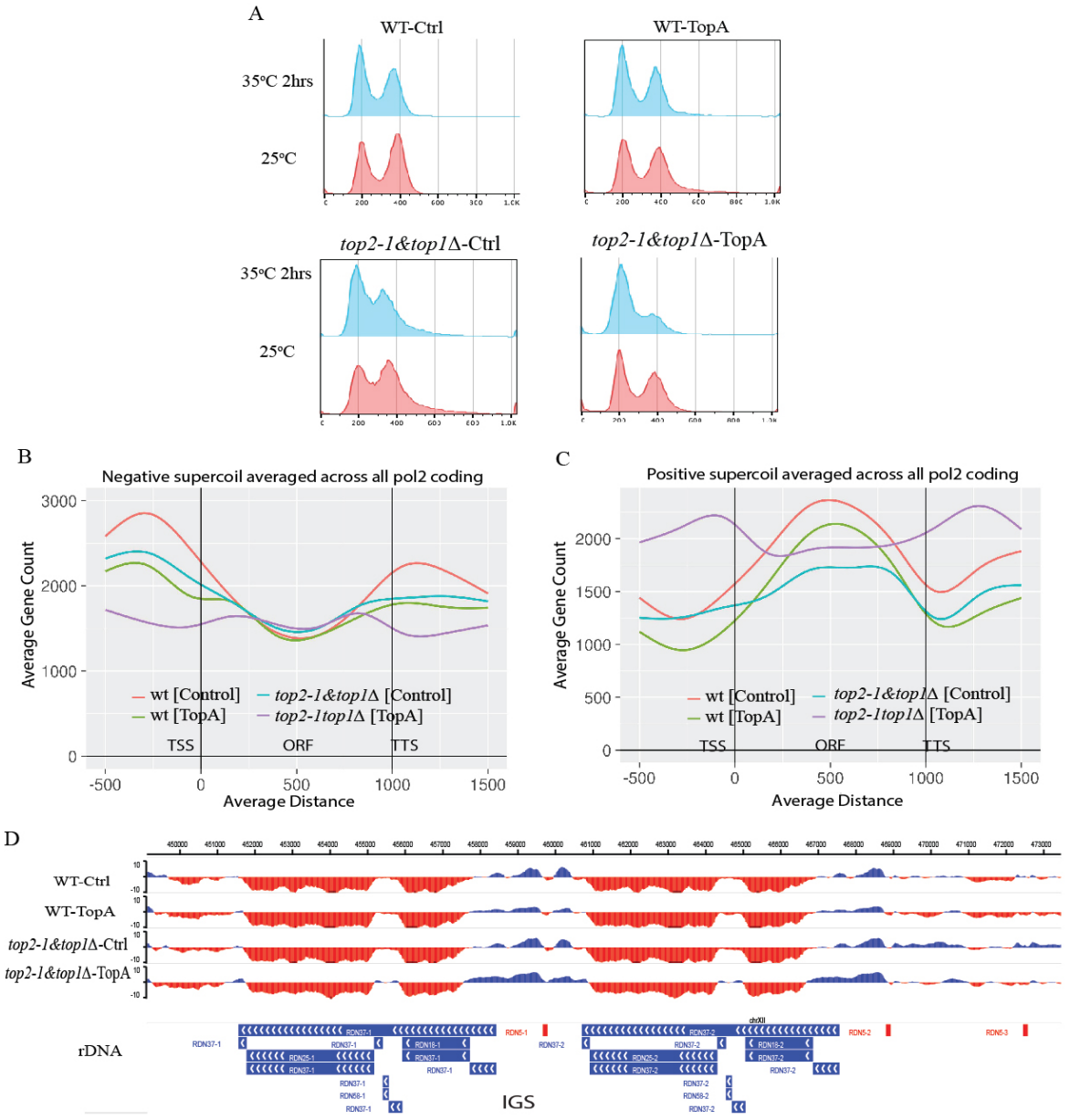


Figure 13: Negative supercoil distribution in the absence of topoisomerases double mutant and expression of E.coli TopA A) FACS profile of four strains WT-Ctrl, WT-TopA, *top2-1&top1*Δ-Ctrl, *top2-1&top1*Δ-TopA before and after moving to restrictive temperature B) MetaORF gene count plot of negative supercoil binding profile averaged across all the ORFs for 1000 bases along with upstream and downstream 500 bases C) positive supercoil binding profile averaged across all the ORFs for 1000 bases along with upstream and downstream 500 bases D) Negative supercoil distribution in the rDNA region (ChrXII:450-490KB), where rDNA section contains two rDNA units with intergenic spacer (IGS) but the normal wildtype contains ~150 copies of rDNA units.

TopA expression in wild type cells showed a reduction of negative supercoil at ORF flanking regions. In *top1Δ top2-1* cells, TopA expression showed significant abolition of negative supercoil at the flanking regions (Figure 13B). Subsequently, there is a significant increase of positive supercoil in gene boundaries in topoisomerase double mutant with TopA expression (Figure 13C), which was not observed in any other genotype. *E.coli* TopA is different from eukaryotic type I topoisomerases, as it specifically acts on negative supercoil converting it into relaxed state (Lee and Garrard 1991). In rDNA, no change was observed in negative or positive supercoil accumulation in all four samples (Figure 13D). Accordingly, the disappearance of negative supercoils at flanking regions paralleled the progressive accumulation of overwound DNA at the same location. TopA expression caused a reduction of positive supercoil at the transcribed regions in wild type cells, while it had the opposite effect in *top1Δ top2*-1 cells. From this observation, we can assume that the diffusion of positive supercoil waves across the entire gene bodies in *top1Δ top2*-1 cells with TopA expression is due to the absence of the negative supercoil cruciform confinements and Top2 protein around the gene bodies.

4.3 Role of topoisomerases and negative supercoil in higher order chromatin organization

4.3.1 Ablation of negative supercoil and inducing positive supercoil disrupts nucleosome binding

The topology of the chromatin is affected by nucleosome positioning (Prunell 1998). We therefore analyzed the nucleosome distribution across the transcribed genes by visualizing the distribution of Histone H3 and investigated whether nucleosome formation was sensitive to supercoil alterations (Lee and Garrard 1991). Histone H3 ChIP-seq revealed that, in wild type cells, H3 was distributed at the transcribed units but was less abundant at the gene boundaries. Wildtype cells having *E.coli* TopA expression did not alter nucleosome

positioning and distribution (Figure 10A-B). Also, in *top1Δ top2-1* mutants, the nucleosome occupancy is similar to wildtype cells. However, in *top1Δ top2-1* mutants with TopA expression a significant reduction (50% reduction compared with control expressed *top1Δ top2-1* mutants) of H3 distribution was seen (Figure 14A-B). Moreover, H3 redistributed its levels to have similar levels at flanking regions and decreased levels at the transcribed units,

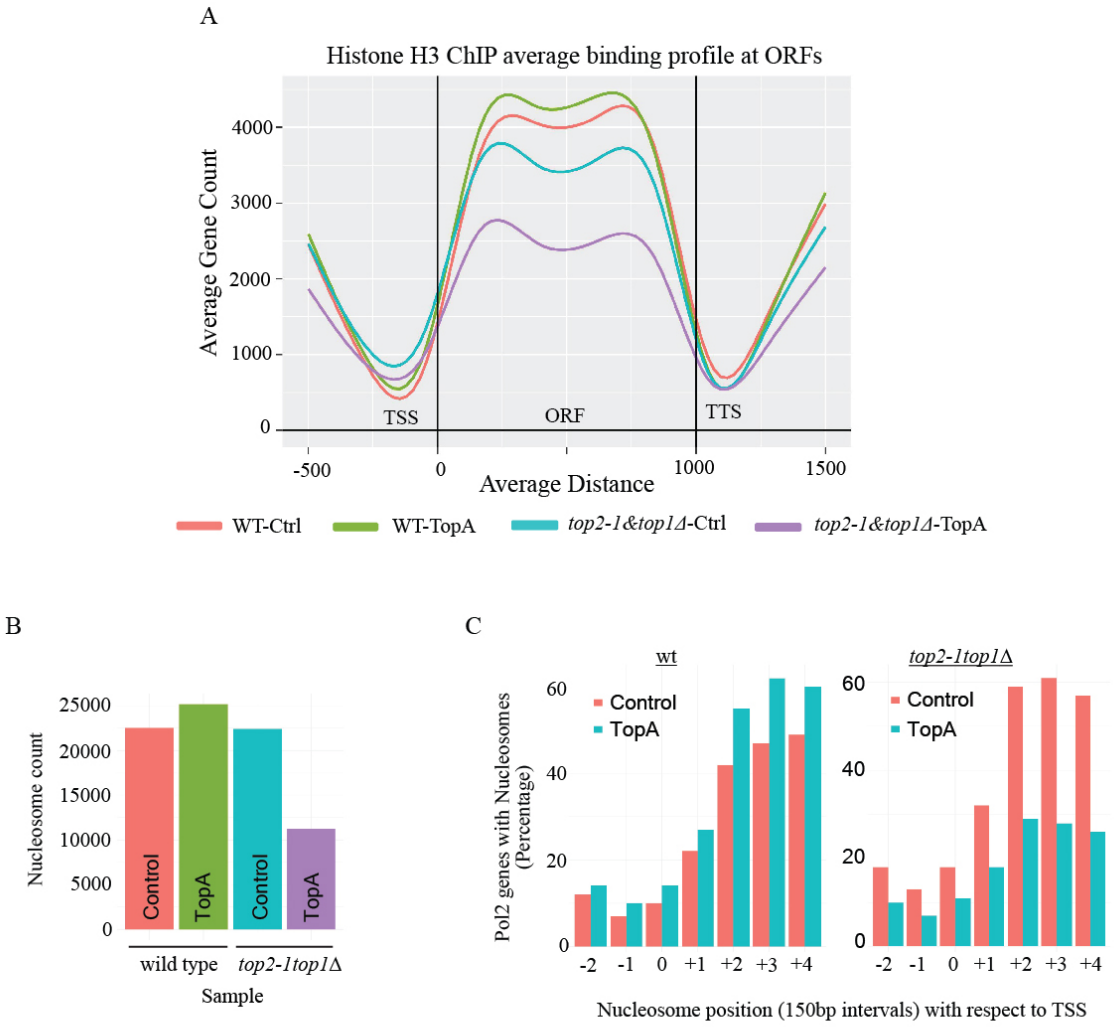


Figure 14: Nucleosome organization in wildtype and topoisomerases double mutant with and without E.coli TopA A) MetaORF gene count plot of Histone (H3) binding averaged across all the ORFs of four strains WT-Ctrl, WT-TopA, *top2top1Δ*-Ctrl and *top2top1Δ*-TopA B) Number of nucleosome peaks (149bp) identified across the four strains C) Percentage of nucleosome peaks identified with respect to transcirption starrt site across the four strains.

starting from position +1 (Figure 14C). The TopA expression in topoisomerase double mutant depletes negative supercoil, followed by diffusion of positive supercoil across the ORF with a corresponding significant reduction in nucleosome occupancy. The formation of nucleosome core is not dependent on the presence or absence of negative supercoil as

suggested by in-vitro studies (Patterton and von Holt 1993). However, upon decreasing the negative torsion, opening of the nucleosome (mean angle opening) was observed (Elbel and Langowski 2015), suggesting that increased positive supercoil destabilizes nucleosomes. Interestingly, DNA supercoil waves can transmit through the nucleosome array, without being significantly blocked by the nucleosome core particles, as demonstrated by single molecule experiments (Lavelle, Victor et al. 2010). Hence, TopA expression in *top1Δ top2-1* cells caused an increase of positive supercoil waves across the entire gene body and significant nucleosome repositioning.

4.3.2 Chromatin conformation capture (Hi-C) to map the three-dimensional organization in yeast

The chromosomes inside the nucleus are organized into multiple levels of hierarchical organization. The higher organization influences various biological processes like DNA repair, gene expression, cell cycle and replication. Perturbation in organization may lead to irregular chromosome condensation during cell cycle and chromosomal abnormalities. Genome wide organization can be studied by chromatin conformation capture techniques like Hi-C. By leveraging the chromatin conformation capture technique, followed by high throughput sequencing (Hi-C), genome wide inter- and intra-chromosomal contact frequencies were calculated to study the chromatin organization for wild-type and topoisomerase double mutant harboring *E.coli* TopA expression. The Hi-C libraries were generated using the four-base cutter DpnII (GATC), followed by ligation and paired end high throughput sequencing. The number of intra (±20 Million) and inter chromosomal interactions (±15 Million) are similar between the wild type cells with control vector and TopA plasmid (Figure 15A).

To visualize the genome organization, the genome was segmented at 5kb resolution to calculate the contact frequencies between all possible pairs of segments and it was converted into two-dimensional (2D) heatmaps which represent the average chromatin

contact maps over a population of cells. In wild-type cells based on the contact frequency, the centromeres of all chromosomes cluster together as shown in the 2D heatmap (Figure 15B & 15C) which indicates that the centromeres are in close proximity with each other in the nucleus. The contact frequency between the two arms (short and long) of individual chromosomes were very low compared to the contact frequency within the same arm of the

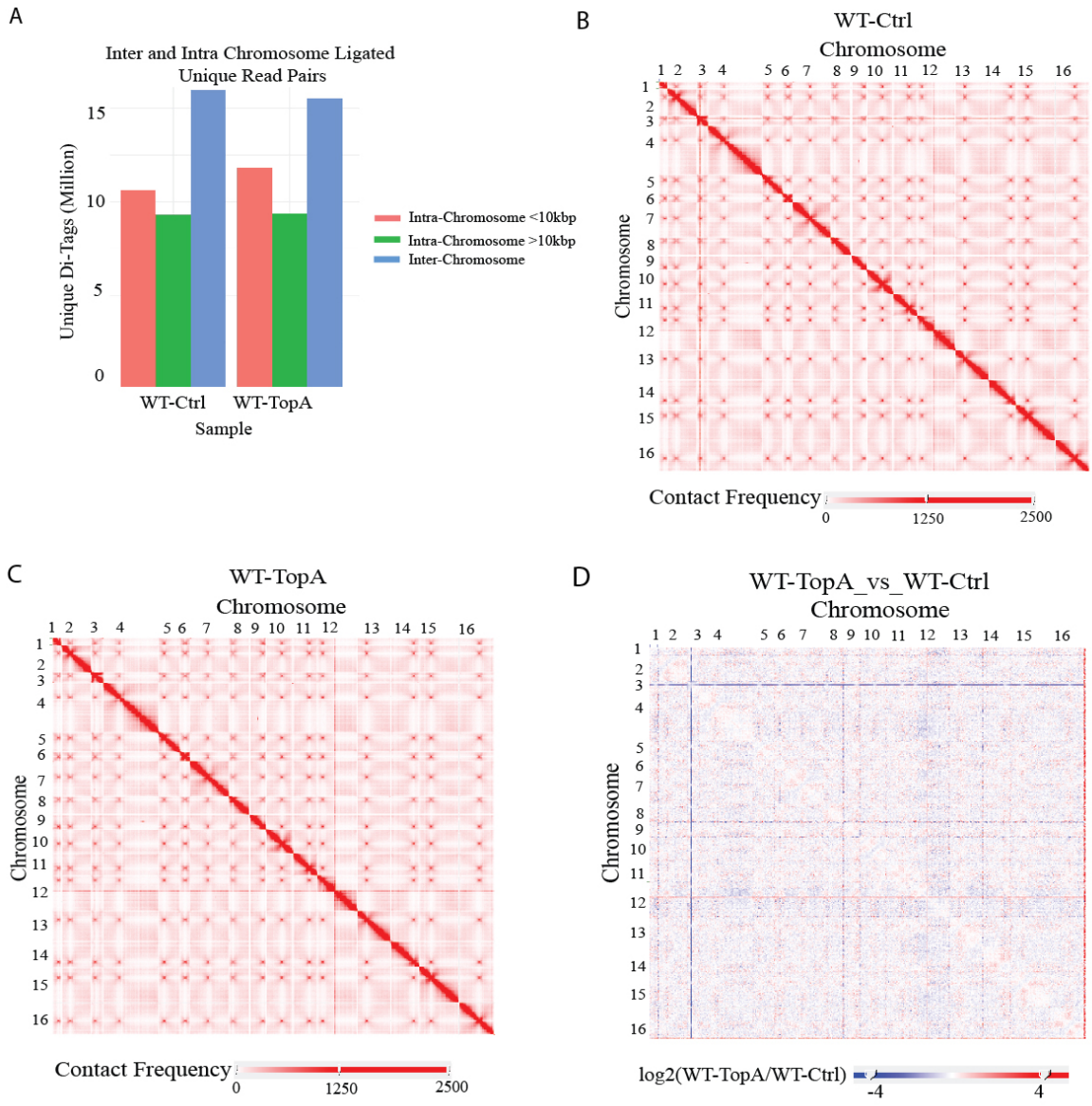


Figure 15: Chromatin organization in wildtype with and without E.coli TopA A) Number of Unique interaction obtained from HiC experiment showing inter- and intra-chromosome interactions. B) HiC heatmap of WT-Ctrl showing the contactact frequency of inter- and intra-chromosome interactions. C) HiC heatmap of WT-TopA showing the contactact frequency of inter- and intra-chromosome interactions. D) log2 differential heatmap between WT-TopA vs WT-Ctrl

chromosomes. The telomeres predominantly interact with other sets of telomeres representing a specific set of telomeres forming clusters. The signal of telomere clusters is

not as robust as the centromere clusters suggesting that there are multiple clusters of telomeres in various regions. The rDNA region interacts with all the other regions of the genome except close to centromere region. These observations suggest that centromeres and groups of telomeres are at opposite poles of the nucleus, and chromosome arms forming distinct territories are in concordance with the previously reported observation of Rabl configuration. We observed no difference in contact frequency between wild type cells with and without TopA plasmid. To measure the difference quantitatively the log ratio between two maps with 5kb resolution is calculated, log ratio for most of the pairwise bins are close to zero and the two-color scale represents the variation in contact frequency between the two conditions (Figure 15D). This shows that the expression of *E.coli* TopA does not alter the genome organization in a significant way.

4.3.2 Impaired chromosome compaction in negative supercoil depleted cells

In *top2-1top1Δ* compared with wildtype cells, the number of intra chromosomal interactions are reduced by ±30% and inter-chromosomal interactions are reduced by ±55% (Figure 16A). In *top2-1top1Δ* with TopA expression compared with wildtype cells, there is no change in the number of intra chromosomal interactions, however there is a significant increase in short-range interactions (<75kb) and decrease in large range interactions (>150kb), also the inter-chromosomal interactions are reduced by ±70%. The contact probability plot (Figure 16B) shows that the topoisomerase *top2-1top1Δ* double mutant with TopA expression has more long range intra-chromosomal interactions (>10kb) compared with the wildtype, whereas in the topoisomerase *top2-1top1Δ* double mutant there is a decrease in the overall intra-chromosomal interactions compared with the wildtype. The wild type with and without TopA expression shows no difference in the contact probability with respect to distance. The enrichment in long-intra chromosomal interactions in topoisomerase *top2-1top1Δ* double mutant with TopA may impair the Rabl configuration of

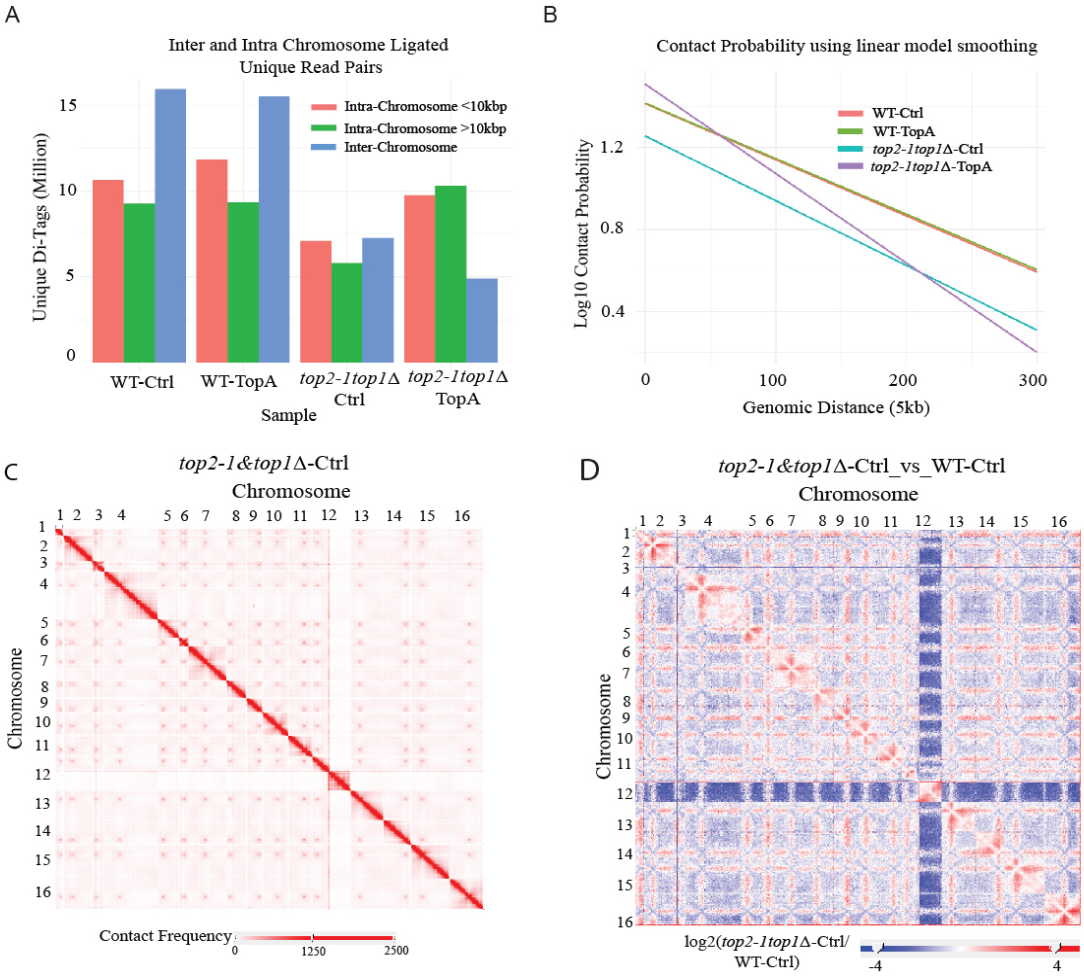


Figure 16: Chromatin organization in the absence of topoisomerases double mutant A) Number of Unique interaction obtained from HiC experiment showing inter- and intra-chromosome interactions. B) Intrachromosome contact probability with respect to geneomic distance (5kb resolution). C) HiC heat-map of *top2-1&top1Δ*-Ctrl showing the contactact frequency of inter- and intra-chromosome interactions. D) log2 differential heatmap between *top2-1&top1Δ*-Ctrl vs WT-Ctrl.

the genome organization where the centromeres cluster together at the spindle pole body and telomeres cluster in groups around the nuclear membrane.

The 2D heat map of *top1Δ top2-1* cells show that, the inter-chromosomal interactions in the right arm of chromosome XII are severely depleted (Figure 16C), where the rDNA repeated units are located in budding yeast. It is been shown that in the absence of topoisomerase double mutant, the rDNA units are highly unstable and forming extra chromosomal rings and as a consequence the number of rDNA units integrated in the chromosome XII significantly reduces (Kim and Wang 1989). The differential heat map between wildtype and topoisomerase double mutant shows that in the rDNA containing

chromosome XII arm interactions are severely depleted and centromeres interactions are marginally preserved. It was shown that centromere interactions are gradually increased during the cell cycle process, where G2M contain higher centromere-centromere interaction for chromatin compaction (Lazar-Stefanita, Scolari et al. 2017). The *top1Δ&top2-1* cells are unbudded cells with 1C DNA content (Figure 13A). The inter-chromosomal interactions inside the nucleus signify the presence of chromosomes close to each other. It was shown that chromosome I interacts preferentially with chromosome XIV (Duan, Andronescu et al. 2010). The loss of inter-chromosomal and gain of intrachromosomal centromere interactions could lead to the loss of chromosome organization where the Rabl configuration is affected and lead to difficulty in chromosome duplication and segregation. The differential heat map of *top1Δ top2-1* compared with wildtype shows the loss of inter and intra chromosomal interactions in the rDNA containing chromosome arm and the gain of intra chromosome interaction in centromeres (Figure 16D).

In topoisomerase double mutant with TopA expression the intra-chromosomal interactions are increased particularly near the centromeres, as a result of which there is no insulation between the two arms of the chromosomes as seen in wildtype and topoisomerase double mutant (Figure 17A). In these cells, we observe an increase in intra-chromosomal interactions compared with *top2-1&top1Δ* cells. The loss of negative supercoiling and introduction of positive supercoiling by TopA in the topoisomerase double mutant causes the chromatin to interact more frequently in short range (<50kb) and less in the long range (>50kb) compared with the wildtype and *top2-1&top1Δ* cells. The short-range interactions are more prone to interact randomly when there are no defined genome organization such as nucleosomes. This makes the chromatin inaccessible for many architectural proteins such as cohesin and condensin. The differential heat map of *top2-1&top1Δ* with TopA expression compared with wildtype shows the loss of inter chromosomal interactions across the genome and the significant gain of intra chromosomal interactions in centromeres (Figure 17B). The differential heat map of *top2-1&top1Δ* with TopA expression compared with *top2-1&top1Δ*

shows that the rDNA containing chromosome arm interactions are not different compared to other intrachromosomal interactions across the genome (Figure 17C).

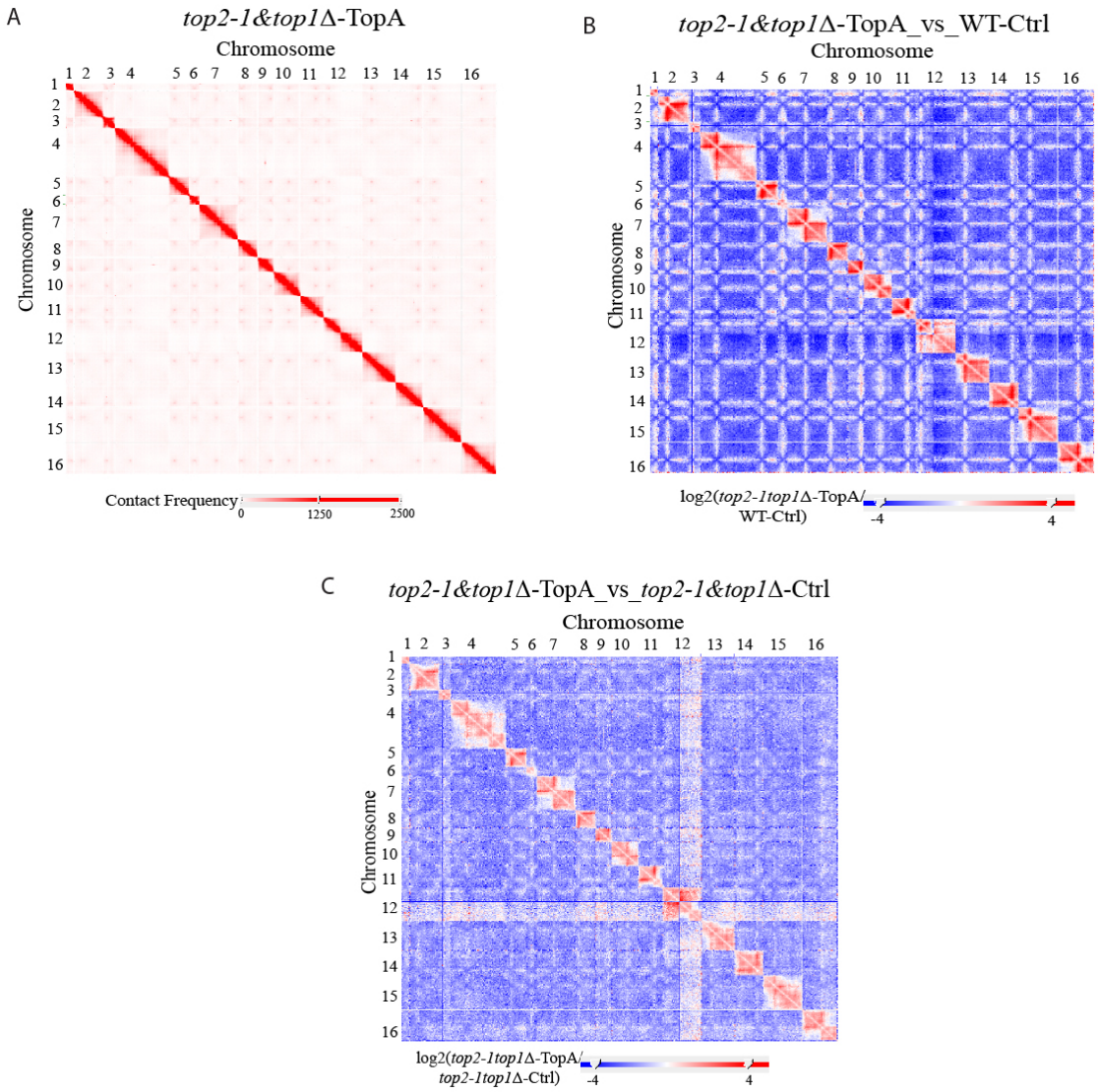


Figure 17: Chromatin organization in the absence of topoisomerases double mutant with expression of E.coli TopA.A) HiC heatmap of *top2-1&top1Δ*-TopA showing the contactact frequency of inter- and intra-chromosome interactions. B) log2 differential heatmap between *top2-1&top1Δ*-TopA vs WT-Ctrl C) log2 differential heatmap between *top2-1&top1Δ*-TopA vs *top2-1&top1Δ*-Ctrl.

4.3.3 Cohesin localization based on transcription and supercoiling

Cohesin, an important architectural protein in topological associated domains formation (TADs), also assists in holding sister chromatids together during S Phase until mitosis. It's been shown that cohesin loads to centromeres and to the promoter regions of pol2 transcribing genes (Ocampo-Hafalla, Munoz et al. 2016). Later, the cohesin translocates from promoter to transcription termination sites. Here, we analyzed the cohesin

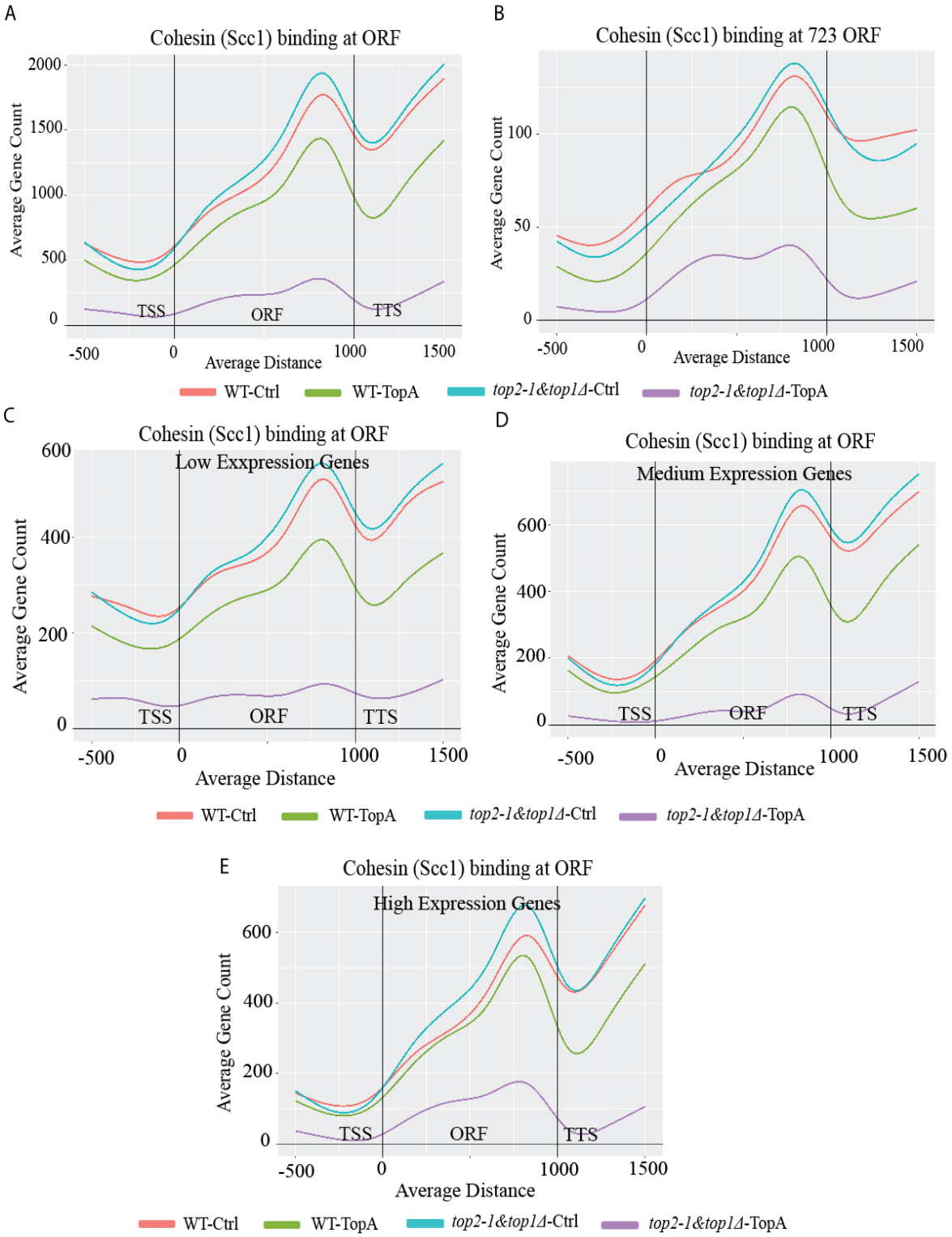


Figure 18: Cohesin protein binding meta-analysis with respect to protein coding genes and transcription A) MetaORF gene count plot of Cohesin (Scc1) binding averaged across all the ORFs of four strains WT-Ctrl, WT-TopA, *top2-1&top1Δ*-Ctrl and *top2-1&top1Δ*-TopA B) MetaORF gene count plot of Cohesin (Scc1) binding averaged across 723 ORFs, where ORFs having no overlapping neighbouring genes C) MetaORF gene count plot of Cohesin (Scc1) binding of four strains WT-Ctrl, WT-TopA, *top2-1&top1Δ*-Ctrl and *top2-1&top1Δ*-TopA averaged across three gene expression classes C) Low D) Medium and E) High gene expressions.

(Scc1 tagged with 10X flag) binding profile in *E.coli* TopA expressed strains under wildtype and *top1Δ top2-1* double mutant. As expected, we saw higher accumulation of cohesin at centromeres and near transcription termination site (TTS) in wild type cells (Figure 18A). There are two peak distributions, one inside the ORF and the other in the intergenic region.

Since the yeast genome is small and many genes are placed close to each other, the second peak could be due to the convergent gene pairs. To test the hypothesis, that the Scc1 protein binding peaks are not from the intergenic regions, the metaorf analysis of protein coding genes having no neighboring genes 500 bases upstream and downstream, showed there are no intergenic peaks proving that the peaks are majorly from the ORF of the protein coding genes (Figure 18B). We analyzed the cohesin accumulation with respect to the gene expression from the RNA-seq, grouping the genes into three class of low, medium and high expression genes. Surprisingly no difference was observed in cohesin binding with respect to low, medium and high expression genes (Figure 18C-D-E). It may be due to the fact that the low rate of transcription is sufficient for cohesin accumulation.

We analyzed the cohesin binding at intergenic spaces with respect to gene orientation by grouping Pol II genes into codirectional (+ and – strands), converging and diverging classes. Converging genes accumulated more cohesin at intergenic spaces, which explains the two-peak distribution near the transcription termination sites (Figure 19A-B-C-D).

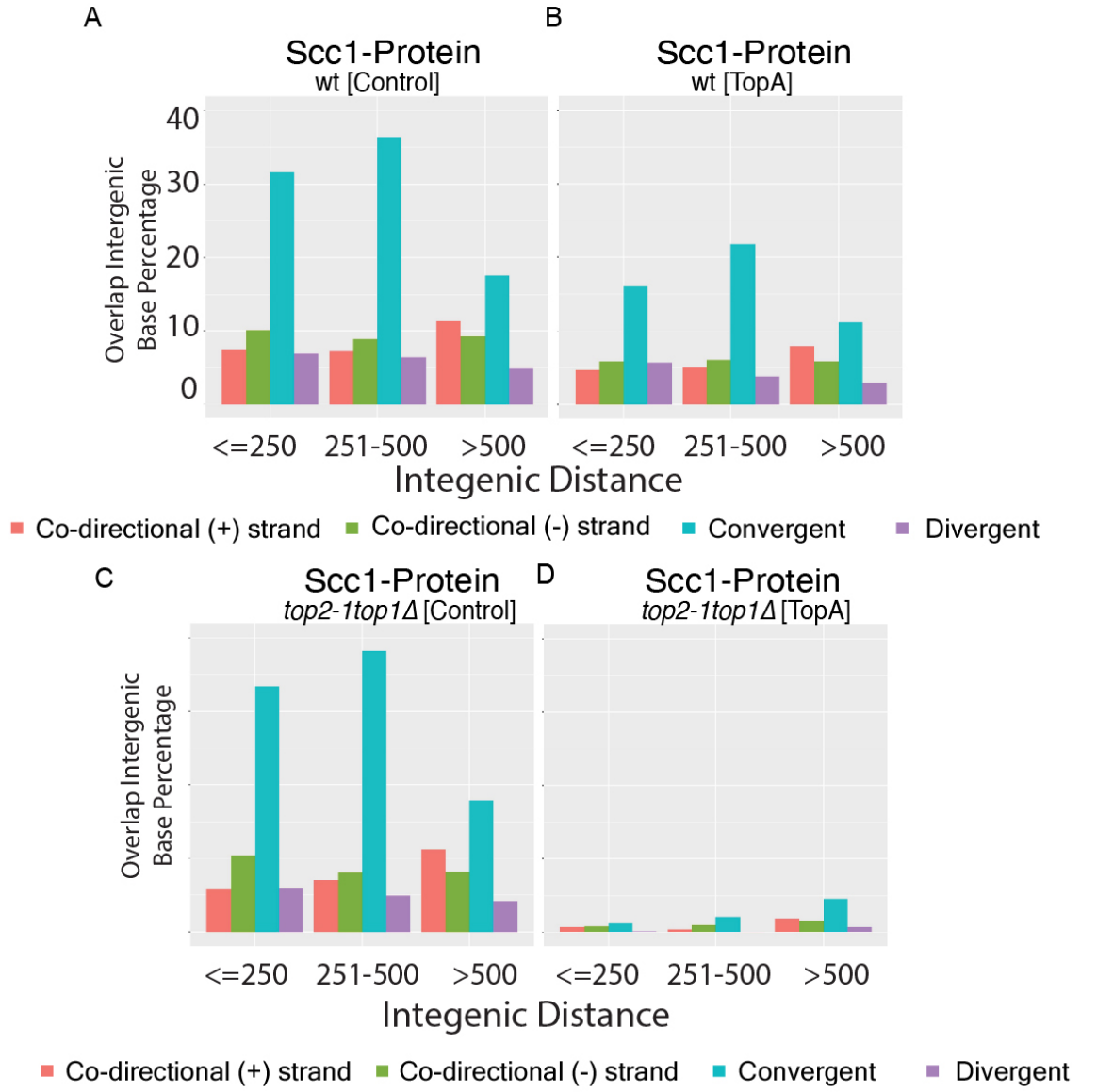


Figure 19: Cohesin distribution based on gene pair transcription directionality. Pol II genes were grouped according to their orientation with respect to neighbouring genes as: Co-directional (+ strand; n=1453 gen pairs), Co-directional (- strand; n=1415 gene pairs), Converging (n=1590 gene pairs) and Diverging (n=1512 gene pairs). Number of gene pairs at different intergenic space (<250 bp = 1729 gene pairs, 251-500bp = 2224 gene pairs and >500 bp = 2010 gene pairs) were plotted with respect to their orientation (Please refer Figure 8 A). A) Base coverage percentage of wt, B) wt[To-pA] C) *top2-1&top1*Δ [Control] D) *top2-1&top1*Δ [TopA] at different intergenic space with respect to gene pairs grouped according to orientation.

The *top2-1&top1Δ* double mutant in ORF shows no changes in Scc1 accumulation compared with wildtype (Figure 18A) but it decreases significantly in the centromeric region (Figure 20A) and no difference was observed in telomeres (Figure 20B). There is a high accumulation of Top2 protein near centromeric and pericentromeric region which helps to trap the cohesin complex near centromere. Top2 is crucial to resolve the cohesin dependent topological stress accumulating in centromeres (Minchell, Keszthelyi et al. 2020).

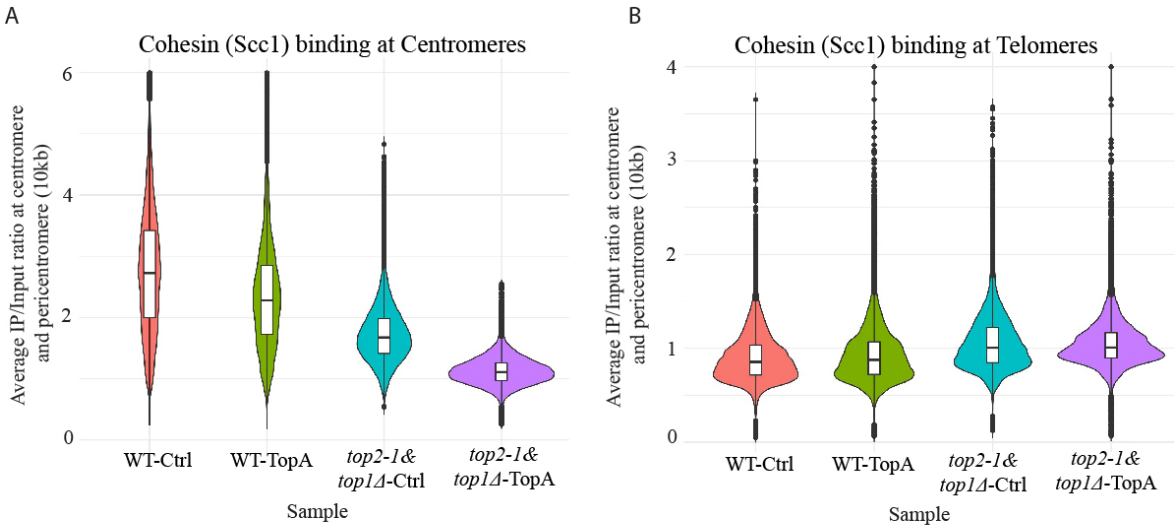


Figure 20: Cohesin protein binding in centromere and telomere A) Boxplot showing average IP/Input of Scc1 binding at centromere and pericentromere (10kb) of four strains WT-Ctrl, WT-TopA, *top2-1&top1Δ*-Ctrl and *top2-1&top1Δ*-TopA. B) Boxplot showing average IP/Input of Scc1 binding at telomeres and sub-telomeres (10kb) of four strains WT-Ctrl, WT-TopA, *top2-1&top1Δ*-Ctrl and *top2-1&top1Δ*-TopA

Wild type and *top2-1&top1Δ* double mutant showed only a slight difference in the number of cohesin binding sites (total number of peaks; WT-Ctrl=1663 & *top2-1&top1Δ*-Ctrl=1896) and peak coverage (median peak width; WT-Ctrl=721 bases & *top2-1&top1Δ*-Ctrl=710 bases). Wild type with and without TopA expression showed no difference in number of binding sites (total number of peaks; WT-Ctrl=1663 & WT-TopA=1805) and peak coverage (median peak width; WT-Ctrl=721 bases & WT-TopA=678 bases), whereas *top2-1&top1Δ* with TopA expression showed significant reduction in both- the number of cohesin binding sites (total number of peaks; WT-TopA=1663 & *top2-1&top1Δ*-TopA=700) as well as peak coverage (median peak width; WT-TopA=721 bases & *top2-1&top1Δ*-TopA=427 bases).

We observe that there are two clusters of peaks based on the peak length: short (median peak width WT=~450 bases) and large peaks (median peak width WT=~1500 bases). The short peaks are prominently seen in promoter region and large peaks are at the end of ORF and centromeres. We speculate that the large peaks represent the permanent residence of cohesion after sliding, whereas the short peaks represent the cohesin loading and sliding. Importantly, the cohesin binding is greatly affected in *top2-1&top1Δ* with TopA

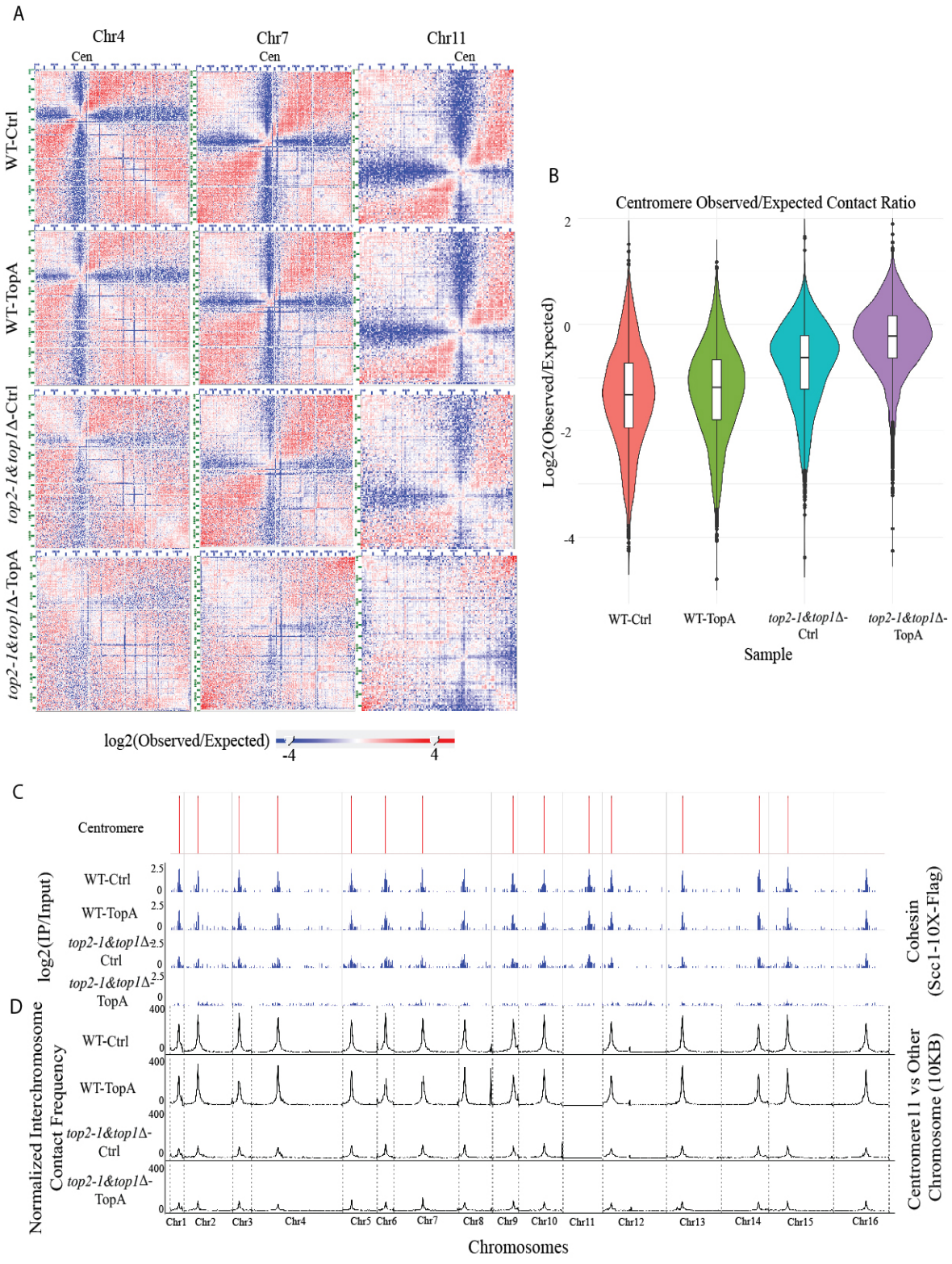


Figure 21: Centromere organization and cohesin protein binding in centromeres A) The Heat-map showing the observed/expected intra-chromosomal interactions along with centromere regions B) Boxplot showing all centromere contact observed/expected ratio of four strains WT-Ctrl, WT-To-pA, *top2-1&top1Δ*-Ctrl and *top2-1&top1Δ*-TopA C) Cohesin binding profile across the whole geneome where all centromeres are highlighted D) Interchromsome interaction of centromere 11 with other chromosomes.

expressed strains. From Hi-C data, the expected value of contact frequency for each chromosome was calculated based on the number of observed interactions and size of the

chromosome bins (5k resolution). The observed/expected value for each chromosome bin is calculated to test for underrepresented and overrepresented regions. The log2(observed/expected) differential heat map for individual chromosomes (Chr4, Chr7, Chr11) shows that centromere organization is affected in *top2-1&top1Δ* double mutant in both Ctrl and TopA plasmid expressed cells (Figure 21A-B) where the absence of Top2 reduces the cohesin accumulation in centromere (Figure 21C).

We conclude that, in *top2-1&top1Δ* double mutant, the cohesin localization in centromeres are affected, altering chromatin organization in centromeres (Figure 21C-D). Modifying the supercoil state at gene boundaries with expressing TopA in *top2-1&top1Δ* cells affect the cohesin localization and chromatin organization across the genome.

5. Discussion

Eukaryotic Top2 and Top1 show functional redundancy, as both can resolve negative and positive supercoiled DNA (Wang 2002). Either one of them is sufficient to ensure progression of DNA replication (Kim and Wang 1989, Bermejo, Doksani et al. 2007) or elongation of the RNAP complex (Mondal and Parvin 2001, Garcia-Rubio and Aguilera 2012, Fernandez, Diaz-Ingelmo et al. 2014, Baranello, Wojtowicz et al. 2016). Here, we show that the action of Top1 and Top2 in S phase is confined to well defined topological domains generated by the architecture of transcribed genes. Top1 accumulation is restricted to ORFs and its activity is influenced by Pol2 levels and gene expression. This is consistent with the finding that Pol2 regulates Top1 activity (Baranello, Wojtowicz et al. 2016); Top2 plays an architectural role in maintaining the negative supercoiled status of ORF flanking regions, to counteract the disruptive potential of incoming replication forks that might dismantle the topological context of transcribed genes, and/or reset the topology of transcribed genes after fork passage.

Our data indicate that the two topoisomerases have non-overlapping functions in normal cells. The ‘twin topological domain model’ (Liu and Wang 1987), predicts that the elongating Pol2 generates negative and positive supercoiling behind and ahead of the transcription bubble, respectively. Negative supercoil influences the rate of formation and stability of transcription bubble at the promoters by weakening base stacking interactions (Revyakin, Liu et al. 2006). However, the presence of negative supercoil was inferred at the beginning and the end of the REP2 gene using a DNase I-based sensitivity assay (Lee and Garrard 1991). The presence of negative supercoil at both sides of a gene would be consistent with a loop-like structure engaging the entire gene body and it is mediated by the interaction between the promoter and the terminator (O'Sullivan, Tan-Wong et al. 2004). Top2 also engages genes in loop-like structures where the promoters and terminators are placed in close proximity(O'Sullivan, Tan-Wong et al. 2004). Based on the gene loops, it is possible

to form multiple twin topological domains by waves of Pol II complexes(Levens, Baranello et al. 2016). In prokaryotes during transcription, the negative supercoil generated from the first transcription bubble adsorb the positive supercoil from the next approaching Pol II complex(Rovinskiy, Agbleke et al. 2012), whereas in eukaryotes, nucleosome assembly might take place following Polymerase II passage (Patterton and von Holt 1993). We showed that, Top1 protein binding at ORF region depends on transcription levels. Based on the observation, RNA Polymerase II elongation might strongly depend on Top1 to resolve topological stress in front of RNA Polymerase II. Polymerase II ORF regions contains a positive supercoiled context, even when transcription is absent implying that genes retain a "memory" of the topological architecture (Liu and Wang 1987, Ma, Bai et al. 2013).

Our data supports the gene loop model as we observed bTMP binding at both sides of the ORFs; notably, the maximum accumulation is upstream of the TSS (around -250 bp) where promoter opening and assembly of the "open complex" takes place (Nagalakshmi, Wang et al. 2008, Erb and van Nimwegen 2011). Out of 3730 genes accumulating positive supercoil within ORFs, 83.83% accumulate Top2 and/or negative supercoil at the ORF flanking regions. Negative supercoil at either sides of a gene would absorb the positive supercoil thus contributing to maintain the chromatin in a decondensed state to facilitate RNA polymerase movement (Wang, Maharana et al. 2014). The gross level of chromatin organization, as well as the degree and magnitude of supercoiling seems maintained throughout the cell cycle (Warren and Cook 1978), and we found that bTMP binding in G1- and S-phase shows 80.1% and 83.5% base coverage similarities in negative and positive supercoil, respectively. Moreover, changes in the topological architecture were not observed when gene expression was repressed, in cell cycle specific genes or genes selectively expressed in certain metabolic conditions, thus implying that the topological context of transcribed genes retains a memory and is not dependent on the local transcription *per se*. It has been shown that negative supercoil accumulates at certain promoters by the transcription-driven supercoil from a nearby promoter (Dunaway and Ostrander 1993). In

this view, it is possible that the initial transcription events may generate waves of torsional forces resulting in the formation of distal topological domains facilitating transcription efficiency. In fact, at a global level, DNA supercoil is dependent on transcription, as negative supercoil is drastically reduced when transcription is inhibited (Kouzine, Gupta et al. 2013, Naughton, Avlonitis et al. 2013).

Majority of the Pol3 transcribed loci, which are responsible for the synthesis of small, conserved non-coding RNAs, such as tRNA, snRNA and snoRNA, were negatively supercoiled in our analysis. Hence, Pol3 transcription seems peculiar, as it can efficiently deal with a negative supercoil context. However, tRNA genes, that are scattered across the genome, often gather at or near the nucleolus (Thompson, Haeusler et al. 2003) in a condensin dependent manner (Gard, Light et al. 2009). We speculate that the peculiar negative supercoil context of Pol3 transcribed loci, might facilitate certain condensation events, which would enhance co-compartmentalization of tRNAs with ribosomal synthesis. Unlike other RNA producing genes, rDNA shows a unique topological context. In line with previous reports (Schultz, Brill et al. 1992, French, Sikes et al. 2011) rRNA genes have high negative supercoil accumulation at the promoters, and positive supercoil at the sequences corresponding to rRNA regions: stable regions were absent. The synthesis of rRNA is strongly inhibited in the absence of Top1 and Top2 (Brill, DiNardo et al. 1987), as the two topoisomerases allow Pol1 to initiate from negative supercoil templates and to overcome the inhibitory effect of positive supercoil during elongation (Schultz, Brill et al. 1992). Additionally, rDNA genes are excised in the form of extrachromosomal rings containing one or more copies of rDNA units in *top1 top2* double mutants (Kim and Wang 1989), a process triggered by negative supercoil-induced recombination (Trigueros and Roca 2002). Accordingly, in our analysis, negative supercoil at the rDNA promoter is increased by 10% in *top1 top2* double mutants, compared to *top2* alone (data not shown).

We propose that the negatively supercoiled and nucleosome free regions that flank the ORFs undergo the formation of DNA cruciform structures characterized by two B-DNA

duplex arms and two intra-strand plectonemic arms in a non-B DNA conformation. Such structures have been visualized and can branch migrate modulating the extension of the intra-strand plectonemic duplexes (Murchie and Lilley 1987, Shlyakhtenko, Potaman et al. 1998, Brazda, Laister et al. 2011). The non-B DNA plectonemic arms counteract nucleosome formation (Nickol and Martin 1983, Nobile, Nickol et al. 1986). HMG box proteins bind to four-ways junctions in their open conformation with high affinity (JR, Norman et al. 1998); Hmo1, a member of the HMG-box protein family, has a preferred binding to four-ways junctions and stabilizes nucleosome free regions and dimerizes, promoting DNA bridging and looping (Kamau, Bauerle et al. 2004, Murugesapillai, McCauley et al. 2014, Panday and Grove 2017). Hmo1 locks cruciform DNA counteracting its branch migration and nucleosome formation. Top2, that colocalizes with Hmo1 (Bermejo, Capra et al. 2009), would act at the base of the loops coordinating their topological transactions with incoming replication forks. The multiple waves of transcription within the same gene loop would generate positive supercoiled domains whose resolution would be mediated by Top1 associated with the Pol2 machinery. Hence, while Top2 acts at the base of the loops, Top1 action would be restricted within the ORF and Hmo1 would contribute to confine the transcription within the gene loops (Bermejo, Capra et al. 2009). We observe chromatin loop formation across all the chromosomes mediated by Top2 protein using ChIA-PET a method that Top2 at the base of the loop helps to counteract it by resolving the topological stress of the incoming fork and to reset the gene topology after fork passage. Top2 mediated chromatin loops contains more than one protein coding genes and can be organized into clusters of chromatin loops where the RNA polymerase clustering could occur in the nucleus (Jackson, Hassan et al. 1993).

In *top2* defective cells negative supercoil decreases at ORF flanking regions and destabilizes the loop formations. This might be due to Top1 activity that is able to convert negative supercoil to stable regions (Koster, Croquette et al. 2005) and is consistent with the finding that, in *top2* mutants, Top1 protein levels dramatically increase at transcribed

genes, particularly at gene boundaries. Notably, non-B DNA structures, can be a substrate for Top1 (Husain, Begum et al. 2016). In *top2* mutants the unrestricted Top1 activity might cause genotoxic events at the non B-DNA cruciform like structures, by extensive nicking and/or knotting (Brown and Cozzarelli 1981). The absence of both Top1 and Top2 affects fork integrity and activates DNA damage check point (Bermejo, Doksani et al. 2007). In topoisomerase double mutant, the negative supercoil accumulation is similar to *top2* mutant. In spite of lack of Top1, the supercoil substrates are destabilized which could be due to DNA breaks.

We studied the effect of supercoiling on genome architecture is studied by expressing *E.coli* TopA in topoisomerase double mutant. It has been previously shown that in yeast cells lacking topoisomerase I and II, the expression of *E.coli* TopA removes the negative supercoil (Gartenberg and Wang 1992). Because the transcription generates both negative and positive supercoil, the removal of negative supercoil by *E.coli* TopA will lead to net accumulation of positive supercoil (Liu and Wang 1987). Based on the bTMP experiment, in topoisomerase double mutant with *E. coli* TopA, negative supercoil are significantly abated in the gene boundaries compared with wildtype expressing TopA. We investigated the chromosome folding and organization with respect to supercoil changes in these yeast strains. We performed Hi-C approach for four genotypes, wildtype and *top2-1&top1Δ* mutants, with and without *E.coli* TopA to obtain unique ligated read pairs and converted them into 5-kb resolution contact maps or heat maps which depicts the average 3D genome organization. We observed the inter chromosomal interactions mostly occur in the centromeres interacting with centromeres of other chromosomes in the wildtype cells and we failed to observe any significant structural changes with the expression of TopA in wildtype cells. The structural differences in chromosome are mostly observed in centromere during the cell cycle, where the establishment of sister chromatid cohesin during replication and condensin-dependent rDNA-centromere clustering during anaphase occurs (Lazar-Stefanita, Scolari et al. 2017).

We observed that the larger chromosome arms are less interacting with other chromosomes compared to the smaller chromosome arms, where the larger chromosomes occupy a distinct territory inside the nucleus. Each chromosome arm forms a domain like structure where there are less interactions between the two arms of the same chromosome compared with the number of interactions in the same arm (Duan, Andronescu et al. 2010). The rDNA which consists of multiple repeats (±150-200 repeats) tends to interact with multiple regions of the chromosomes but doesn't interact with the centromeres, places itself opposite to centromeres in nucleus. Combining all these observations into a three-dimensional space results in Rabl configuration of the yeast chromosomes, where the centromeres are tethered to spindle pole body by microtubules, the chromosome arms extend towards the nuclear membrane and rDNA is placed opposite to the spindle pole body (Berger, Cabal et al. 2008).

The attachment of centromere and telomere to the nuclear membrane and the Rabl configuration is crucial to reduce the topological entanglement of DNA molecules (Pouokam, Cruz et al. 2019). In *top2-1top1Δ* mutant, the interaction of centromere with centromeres of other chromosomes is significantly affected compared to the wildtype. The rDNA inter-chromosomal interactions with other non-centromeric regions are also significantly affected compared to the wildtype. It has been shown that, in topoisomerase double mutant the rDNA units are unstable and excised from the chromosome as extrachromosomal rings (Kim and Wang 1989). This explains the absence of rDNA inter-chromosomal interactions in *top2-1top1Δ* mutant.

In *top2-1top1Δ* with TopA plasmid expression, there is a significant loss of inter-chromosomal interactions and increase of intra-chromosomal interactions across the chromosome and prominently around the centromeres which represent the absence of insulation between the two arms of the same chromosome. The Rabl configuration is completely affected when the interactions between two arms of the chromosomes are same as the interaction within the same arm, the centromeres don't interact with centromeres of

other chromosomes and the absence of interaction in rDNA region of chromosome XII. The expression of TopA plasmid in *top2-1top1Δ* mutants leads to increased accumulation of positive supercoiling and decreased nucleosome occupancy across the genome which causes the chromatin to interact more frequently within the chromosome in the short range and lesser in the long range.

The cohesin complex which mediates the chromosome compaction helps in tethering the centromeres to spindle pole body. The Scc1, a cohesin subunit prominently seen in centromeres in wild type condition, is severely depleted in *top2-1top1Δ* mutant with TopA expression. Other than centromeres, the Scc1 binding is observed close to the transcription termination regions of the PolII genes and the accumulation of Scc1 is not completely dependent on the transcription. The cohesin is loaded onto the promoter of the active genes and slides towards the transcription termination region. It is topologically bound to the DNA and sliding is not limited to transcription (Ocampo-Hafalla, Munoz et al. 2016).

The convergent gene pair orientation accumulates high amount of cohesin compared with divergent and codirectional gene pairs. The transcription induced supercoiling can assist the sliding of cohesin (Racko, Benedetti et al. 2018), whereas the arrest of sliding could be due to the presence of non-B-DNA or secondary structure formed by the negative supercoil in the gene boundary since there are no CTCF like protein in yeast. There is a high amount of Top2 protein accumulation in the pericentromeric region, very close to the Scc1 protein. This helps to preserve the topological context and restricts the Scc1 sliding from the centromere. In *top2-1top1Δ* mutant, Scc1 protein accumulation in centromere is depleted, whereas the other regions are preserved and similar to wildtype condition. In *top2-1top1Δ* mutant with TopA, Scc1 protein is significantly depleted in all the regions including centromeres. Due to the absence of negative supercoiling in gene boundaries of *top2-1top1Δ* mutant with TopA expression, there is a significant reduction of nucleosome occupancy across the genome. The formation of nucleosome core is not dependent on presence or absence of negative supercoil, as was also suggested by in-vitro studies (Patterton and von

Holt 1993). However, upon decreasing the negative torsion effect, the accessibility of nucleosome and the opening of the nucleosome is affected (Elbel and Langowski 2015), suggesting that increased positive supercoil destabilizes nucleosomes.

Based on all the above observation, the Rabl configuration is completely affected when negative supercoiling is excised and net state of the genome is positively supercoiled. This affects the nucleosome occupancy and leads to an increase in short range intrachromosomal interaction and loss of long-range interactions. Due to this phenomenon, the genome poses more topological entanglement and affects the binding of chromatin architectural proteins such as cohesin.

In the future, we plan to extend the bTMP experiment to the mammalian system and study the role of DNA supercoil in the higher order genome organization. We also plan to visualize the DNA cruciform structures using DNA electron microscopy.

6. References

Anderson, D. E., A. Losada, H. P. Erickson and T. Hirano (2002). "Condensin and cohesin display different arm conformations with characteristic hinge angles." J Cell Biol **156**(3): 419-424.

Ansari, A. and M. Hampsey (2005). "A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping." Genes Dev **19**(24): 2969-2978.

Apostolou, E., F. Ferrari, R. M. Walsh, O. Bar-Nur, M. Stadtfeld, S. Cheloufi, H. T. Stuart, J. M. Polo, T. K. Ohsumi, M. L. Borowsky, P. V. Kharchenko, P. J. Park and K. Hochedlinger (2013). "Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming." Cell Stem Cell **12**(6): 699-712.

Arlt, M. F. and T. W. Glover (2010). "Inhibition of topoisomerase I prevents chromosome breakage at common fragile sites." DNA Repair (Amst) **9**(6): 678-689.

Austin, C. A., J. H. Sng, S. Patel and L. M. Fisher (1993). "Novel HeLa topoisomerase II is the II beta isoform: complete coding sequence and homology with other type II topoisomerases." Biochim Biophys Acta **1172**(3): 283-291.

Baranello, L., D. Wojtowicz, K. Cui, B. N. Devaiah, H. J. Chung, K. Y. Chan-Salis, R. Guha, K. Wilson, X. Zhang, H. Zhang, J. Piotrowski, C. J. Thomas, D. S. Singer, B. F. Pugh, Y. Pommier, T. M. Przytycka, F. Kouzine, B. A. Lewis, K. Zhao and D. Levens (2016). "RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription." Cell **165**(2): 357-371.

Baxter, J. and J. F. Diffley (2008). "Topoisomerase II inactivation prevents the completion of DNA replication in budding yeast." Mol Cell **30**(6): 790-802.

Berger, A. B., G. G. Cabal, E. Fabre, T. Duong, H. Buc, U. Nehrbass, J. C. Olivo-Marin, O. Gadal and C. Zimmer (2008). "High-resolution statistical mapping reveals gene territories in live yeast." Nat Methods **5**(12): 1031-1037.

Bermejo, R., T. Capra, V. Gonzalez-Huici, D. Fachinetti, A. Cocito, G. Natoli, Y. Katou, H. Mori, K. Kurokawa, K. Shirahige and M. Foiani (2009). "Genome-organizing factors Top2 and Hmo1 prevent chromosome fragility at sites of S phase transcription." Cell **138**(5): 870-884.

Bermejo, R., Y. Doksani, T. Capra, Y. M. Katou, H. Tanaka, K. Shirahige and M. Foiani (2007). "Top1- and Top2-mediated topological transitions at replication forks ensure fork progression and stability and prevent DNA damage checkpoint activation." Genes Dev **21**(15): 1921-1936.

Bermejo, R., Y. M. Katou, K. Shirahige and M. Foiani (2009). "ChIP-on-chip analysis of DNA topoisomerases." Methods Mol Biol **582**: 103-118.

Bianchi, M. E., M. Beltrame and G. Paonessa (1989). "Specific recognition of cruciform DNA by nuclear protein HMG1." Science **243**(4894 Pt 1): 1056-1059.

Blasquez, V. C., A. O. Sperry, P. N. Cockerill and W. T. Garrard (1989). "Protein:DNA interactions at chromosomal loop attachment sites." Genome **31**(2): 503-509.

Brazda, V., R. C. Laister, E. B. Jagelska and C. Arrowsmith (2011). "Cruciform structures are a common DNA feature important for regulating biological processes." BMC Mol Biol **12**: 33.

Brill, S. J., S. DiNardo, K. Voelkel-Meiman and R. Sternglanz (1987). "Need for DNA topoisomerase activity as a swivel for DNA replication for transcription of ribosomal RNA." Nature **326**(6111): 414-416.

Brown, P. O. and N. R. Cozzarelli (1981). "Catenation and knotting of duplex DNA by type 1 topoisomerases: a mechanistic parallel with type 2 topoisomerases." Proc Natl Acad Sci U S A **78**(2): 843-847.

Bystricky, K., T. Laroche, G. van Houwe, M. Blaszczyk and S. M. Gasser (2005). "Chromosome looping in yeast: telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization." J Cell Biol **168**(3): 375-387.

Campbell, J. L. and O. Cohen-Fix (2002). "Chromosome cohesion: ring around the sisters?" Trends Biochem Sci **27**(10): 492-495.

Capranico, G., C. Jaxel, M. Roberge, K. W. Kohn and Y. Pommier (1990). "Nucleosome positioning as a critical determinant for the DNA cleavage sites of mammalian DNA topoisomerase II in reconstituted simian virus 40 chromatin." Nucleic Acids Res **18**(15): 4553-4559.

Champoux, J. J. (2001). "DNA topoisomerases: structure, function, and mechanism." Annu Rev Biochem **70**: 369-413.

Chang, C. R., C. S. Wu, Y. Hom and M. R. Gartenberg (2005). "Targeting of cohesin by transcriptionally silent chromatin." Genes Dev **19**(24): 3031-3042.

Chen, W. H., W. Wei and M. J. Lercher (2011). "Minimal regulatory spaces in yeast genomes." BMC Genomics **12**: 320.

Christman, M. F., F. S. Dietrich and G. R. Fink (1988). "Mitotic recombination in the rDNA of S. cerevisiae is suppressed by the combined action of DNA topoisomerases I and II." Cell **55**(3): 413-425.

Ciesla, Z., K. Mardarowicz and T. Klopotowski (1974). "Inhibition of DNA synthesis and cell division in Salmonella typhimurium by azide." Mol Gen Genet **135**(4): 339-348.

Cole, H. A., B. H. Howard and D. J. Clark (2011). "The centromeric nucleosome of budding yeast is perfectly positioned and covers the entire centromere." Proc Natl Acad Sci U S A **108**(31): 12687-12692.

Cotta-Ramusino, C., D. Fachinetti, C. Lucca, Y. Doksani, M. Lopes, J. Sogo and M. Foiani (2005). "Exo1 processes stalled replication forks and counteracts fork reversal in checkpoint-defective cells." Mol Cell **17**(1): 153-159.

Cullen, K. E., M. P. Kladde and M. A. Seyfred (1993). "Interaction between transcription regulatory regions of prolactin chromatin." Science **261**(5118): 203-206.

Dekker, J. and E. Heard (2015). "Structural and functional diversity of Topologically Associating Domains." FEBS Lett **589**(20 Pt A): 2877-2884.

Dekker, J. and L. Mirny (2016). "The 3D Genome as Moderator of Chromosomal Communication." Cell **164**(6): 1110-1121.

Dekker, J., K. Rippe, M. Dekker and N. Kleckner (2002). "Capturing chromosome conformation." Science **295**(5558): 1306-1311.

Dionne, I. and R. J. Wellinger (1998). "Processing of telomeric DNA ends requires the passage of a replication fork." Nucleic Acids Res **26**(23): 5365-5371.

Dixon, J. R., I. Jung, S. Selvaraj, Y. Shen, J. E. Antosiewicz-Bourget, A. Y. Lee, Z. Ye, A. Kim, N. Rajagopal, W. Xie, Y. Diao, J. Liang, H. Zhao, V. V. Lobanenkov, J. R. Ecker, J. A. Thomson and B. Ren (2015). "Chromatin architecture reorganization during stem cell differentiation." Nature **518**(7539): 331-336.

Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu and B. Ren (2012). "Topological domains in mammalian genomes identified by analysis of chromatin interactions." Nature **485**(7398): 376-380.

Droit, A., C. Cheung and R. Gottardo (2010). "rMAT--an R/Bioconductor package for analyzing ChIP-chip experiments." Bioinformatics **26**(5): 678-679.

Duan, Z., M. Andronescu, K. Schutz, S. McIlwain, Y. J. Kim, C. Lee, J. Shendure, S. Fields, C. A. Blau and W. S. Noble (2010). "A three-dimensional model of the yeast genome." Nature **465**(7296): 363-367.

Dunaway, M. and E. A. Ostrander (1993). "Local Domains of Supercoiling Activate a Eukaryotic Promoter Invivo." Nature **361**(6414): 746-748.

Elbel, T. and J. Langowski (2015). "The effect of DNA supercoiling on nucleosome structure and stability." J Phys Condens Matter **27**(6): 064105.

Erb, I. and E. van Nimwegen (2011). "Transcription Factor Binding Site Positioning in Yeast: Proximal Promoter Motifs Characterize TATA-Less Promoters." Plos One **6**(9).

Fachinetti, D., R. Bermejo, A. Cocito, S. Minardi, Y. Katou, Y. Kanoh, K. Shirahige, A. Azvolinsky, V. A. Zakian and M. Foiani (2010). "Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements." Mol Cell **39**(4): 595-605.

Fernandez, X., O. Diaz-Ingelmo, B. Martinez-Garcia and J. Roca (2014). "Chromatin regulates DNA torsional energy via topoisomerase II-mediated relaxation of positive supercoils." EMBO J **33**(13): 1492-1501.

Franke, M., D. M. Ibrahim, G. Andrey, W. Schwarzer, V. Heinrich, R. Schopflin, K. Kraft, R. Kempfer, I. Jerkovic, W. L. Chan, M. Spielmann, B. Timmermann, L. Wittler, I. Kurth, P. Cambiaso, O. Zuffardi, G. Houge, L. Lambie, F. Brancati, A. Pombo, M. Vingron, F. Spitz and S. Mundlos (2016). "Formation of new chromatin domains determines pathogenicity of genomic duplications." Nature **538**(7624): 265-269.

French, S. L., M. L. Sikes, R. D. Hontz, Y. N. Osheim, T. E. Lambert, A. El Hage, M. M. Smith, D. Tollervey, J. S. Smith and A. L. Beyer (2011). "Distinguishing the roles of Topoisomerases I and II in relief of transcription-induced torsional stress in yeast rRNA genes." Mol Cell Biol **31**(3): 482-494.

Fudenberg, G., M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur and L. A. Mirny (2016). "Formation of Chromosomal Domains by Loop Extrusion." Cell Rep **15**(9): 2038-2049.

Garcia-Rubio, M. L. and A. Aguilera (2012). "Topological constraints impair RNA polymerase II transcription and causes instability of plasmid-borne convergent genes." Nucleic Acids Research **40**(3): 1050-1064.

Gard, S., W. Light, B. Xiong, T. Bose, A. J. McNairn, B. Harris, B. Fleharty, C. Seidel, J. H. Brickner and J. L. Gerton (2009). "Cohesinopathy mutations disrupt the subnuclear organization of chromatin." J Cell Biol **187**(4): 455-462.

Gartenberg, M. R. and J. C. Wang (1992). "Positive Supercoiling of DNA Greatly Diminishes Messenger-Rna Synthesis in Yeast." Proceedings of the National Academy of Sciences of the United States of America **89**(23): 11461-11465.

Gartenberg, M. R. and J. C. Wang (1992). "Positive supercoiling of DNA greatly diminishes mRNA synthesis in yeast." Proc Natl Acad Sci U S A **89**(23): 11461-11465.

Gassler, J., H. B. Brandao, M. Imakaev, I. M. Flyamer, S. Ladstatter, W. A. Bickmore, J. M. Peters, L. A. Mirny and K. Tachibana (2017). "A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture." EMBO J **36**(24): 3600-3618.

Ghavi-Helm, Y., F. A. Klein, T. Pakozdi, L. Ciglar, D. Noordermeer, W. Huber and E. E. Furlong (2014). "Enhancer loops appear stable during development and are associated with paused polymerase." Nature **512**(7512): 96-100.

Gilmour, D. S. and J. T. Lis (1986). "RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells." Mol Cell Biol **6**(11): 3984-3989.

Glynn, E. F., P. C. Megee, H. G. Yu, C. Mistrot, E. Unal, D. E. Koshland, J. L. DeRisi and J. L. Gerton (2004). "Genome-wide mapping of the cohesin complex in the yeast Saccharomyces cerevisiae." PLoS Biol **2**(9): E259.

Gotta, M., T. Laroche, A. Formenton, L. Maillet, H. Scherthan and S. M. Gasser (1996). "The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type Saccharomyces cerevisiae." J Cell Biol **134**(6): 1349-1363.

Greenfeder, S. A. and C. S. Newlon (1992a). "Replication forks pause at yeast centromeres." Mol Cell Biol **12**(9): 4056-4066.

Greenfeder, S. A. and C. S. Newlon (1992b). "A replication map of a 61-kb circular derivative of Saccharomyces cerevisiae chromosome III." Mol Biol Cell **3**(9): 999-1013.

Gruber, S., C. H. Haering and K. Nasmyth (2003). "Chromosomal cohesin forms a ring." Cell **112**(6): 765-777.

Haering, C. H., J. Lowe, A. Hochwagen and K. Nasmyth (2002). "Molecular architecture of SMC proteins and the yeast cohesin complex." Mol Cell **9**(4): 773-788.

Harvey, J., S. C. Hardy and M. L. Ashford (1999). "Dual actions of the metabolic inhibitor, sodium azide on K(ATP) channel currents in the rat CRI-G1 insulinoma cell line." Br J Pharmacol **126**(1): 51-60.

Hnisz, D., A. S. Weintraub, D. S. Day, A. L. Valton, R. O. Bak, C. H. Li, J. Goldmann, B. R. Lajoie, Z. P. Fan, A. A. Sigova, J. Reddy, D. Borges-Rivera, T. I. Lee, R. Jaenisch, M. H. Porteus, J. Dekker and R. A. Young (2016). "Activation of proto-oncogenes by disruption of chromosome neighborhoods." Science **351**(6280): 1454-1458.

Hsieh, T. H., A. Weiner, B. Lajoie, J. Dekker, N. Friedman and O. J. Rando (2015). "Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C." Cell **162**(1): 108-119.

Husain, A., N. A. Begum, T. Taniguchi, H. Taniguchi, M. Kobayashi and T. Honjo (2016). "Chromatin remodeller SMARCA4 recruits topoisomerase 1 and suppresses transcription-associated genomic instability." Nat Commun **7**: 10549.

Ivanov, D., A. Schleiffer, F. Eisenhaber, K. Mechtler, C. H. Haering and K. Nasmyth (2002). "Eco1 is a novel acetyltransferase that can acetylate proteins involved in cohesion." Curr Biol **12**(4): 323-328.

Jackson, D. A., A. B. Hassan, R. J. Errington and P. R. Cook (1993). "Visualization of focal sites of transcription within human nuclei." EMBO J **12**(3): 1059-1065.

Jin, Q. W., J. Fuchs and J. Loidl (2000). "Centromere clustering is a major determinant of yeast interphase nuclear organization." J Cell Sci **113 (Pt 11)**: 1903-1912.

Joshi, R. S., C. Nikolaou and J. Roca (2018). "Structure and Chromosomal Organization of Yeast Genes Regulated by Topoisomerase II." Int J Mol Sci **19**(1).

Joshi, R. S., B. Pina and J. Roca (2012). "Topoisomerase II is required for the production of long Pol II gene transcripts in yeast." Nucleic Acids Res **40**(16): 7907-7915.

JR, P. o., D. G. Norman, J. Bramham, M. E. Bianchi and D. M. Lilley (1998). "HMG box proteins bind to four-way DNA junctions in their open conformation." EMBO J **17**(3): 817-826.

Kamau, E., K. T. Bauerle and A. Grove (2004). "The Saccharomyces cerevisiae high mobility group box protein HMO1 contains two functional DNA binding domains." J Biol Chem **279**(53): 55234-55240.

Kim, R. A. and J. C. Wang (1989). "Function of DNA topoisomerases as replication swivels in Saccharomyces cerevisiae." J Mol Biol **208**(2): 257-267.

Kim, R. A. and J. C. Wang (1989). "A subthreshold level of DNA topoisomerases leads to the excision of yeast rDNA as extrachromosomal rings." Cell **57**(6): 975-985.

Klobutcher, L. A., M. T. Swanton, P. Donini and D. M. Prescott (1981). "All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus." Proc Natl Acad Sci U S A **78**(5): 3015-3019.

Kobayashi, T. and A. R. Ganley (2005). "Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats." Science **309**(5740): 1581-1584.

Kobayashi, T., D. J. Heck, M. Nomura and T. Horiuchi (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 Dev **12**(24): 3821-3830.

Kobayashi, T., T. Horiuchi, P. Tongaonkar, L. Vu and M. Nomura (2004). "SIR2 regulates recombination between different rDNA repeats, but not recombination within individual rRNA genes in yeast." Cell **117**(4): 441-453.

Koster, D. A., V. Croquette, C. Dekker, S. Shuman and N. H. Dekker (2005). "Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB." Nature **434**(7033): 671-674.

Kouzine, F., A. Gupta, L. Baranello, D. Wojtowicz, K. Ben-Aissa, J. Liu, T. M. Przytycka and D. Levens (2013). "Transcription-dependent dynamic supercoiling is a short-range genomic force." Nat Struct Mol Biol **20**(3): 396-403.

Krijger, P. H. and W. de Laat (2016). "Regulation of disease-associated gene expression in the 3D genome." Nat Rev Mol Cell Biol **17**(12): 771-782.

Kristin Brogaard, L. X., Ji-Ping Wang, and Jonathan Widom (2012). "A map of nucleosome positions in yeast at base-pair resolution." Nature **486**(7404): 496-501.

Krivega, I. and A. Dean (2012). "Enhancer and promoter interactions-long distance calls." Curr Opin Genet Dev **22**(2): 79-85.

Lavelle, C., J. M. Victor and J. Zlatanova (2010). "Chromatin fiber dynamics under tension and torsion." Int J Mol Sci **11**(4): 1557-1579.

Lazar-Stefanita, L., V. F. Scolari, G. Mercy, H. Muller, T. M. Guerin, A. Thierry, J. Mozziconacci and R. Koszul (2017). "Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle." EMBO J **36**(18): 2684-2697.

Lee, M. S. and W. T. Garrard (1991). "Positive DNA supercoiling generates a chromatin conformation characteristic of highly active genes." Proc Natl Acad Sci U S A **88**(21): 9675-9679.

Lengronne, A., Y. Katou, S. Mori, S. Yokobayashi, G. P. Kelly, T. Itoh, Y. Watanabe, K. Shirahige and F. Uhlmann (2004). "Cohesin relocation from sites of chromosomal loading to places of convergent transcription." Nature **430**(6999): 573-578.

Leonard, A. C. and J. E. Grimwade (2010). "Initiation of DNA Replication." EcoSal Plus **4**(1).

Levens, D., L. Baranello and F. Kouzine (2016). "Controlling gene expression by DNA mechanics: emerging insights and challenges." Biophys Rev **8**(3): 259-268.

Li, T. K., A. Y. Chen, C. Yu, Y. Mao, H. Wang and L. F. Liu (1999). "Activation of topoisomerase II-mediated excision of chromosomal DNA loops during oxidative stress." Genes Dev **13**(12): 1553-1560.

Li, X., O. J. Luo, P. Wang, M. Zheng, D. Wang, E. Piecuch, J. J. Zhu, S. Z. Tian, Z. Tang, G. Li and Y. Ruan (2017). "Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions." Nat Protoc **12**(5): 899-915.

Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander and J. Dekker (2009). "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." Science **326**(5950): 289-293.

Linka, R. M., A. C. Porter, A. Volkov, C. Mielke, F. Boege and M. O. Christensen (2007). "C-terminal regions of topoisomerase IIalpha and IIbeta determine isoform-specific functioning of the enzymes in vivo." Nucleic Acids Res **35**(11): 3810-3822.

Liu, L. F. and J. C. Wang (1987). "Supercoiling of the DNA template during transcription." Proc Natl Acad Sci U S A **84**(20): 7024-7027.

Ljungman, M. and P. C. Hanawalt (1995). "Presence of negative torsional tension in the promoter region of the transcriptionally poised dihydrofolate reductase gene in vivo." Nucleic Acids Res **23**(10): 1782-1789.

Ma, J., L. Bai and M. D. Wang (2013). "Transcription under torsion." Science **340**(6140): 1580-1583.

Manuelidis, L. (1985). "Individual interphase chromosome domains revealed by in situ hybridization." Hum Genet **71**(4): 288-293.

Marchand, C., P. Pourquier, G. S. Laco, N. Jing and Y. Pommier (2002). "Interaction of human nuclear topoisomerase I with guanosine quartet-forming and guanosine-rich single-stranded DNA and RNA oligonucleotides." J Biol Chem **277**(11): 8906-8911.

Meczes, E. L., K. L. Marsh, L. M. Fisher, M. P. Rogers and C. A. Austin (1997). "Complementation of temperature-sensitive topoisomerase II mutations in Saccharomyces cerevisiae by a human TOP2 beta construct allows the study of topoisomerase II beta inhibitors in yeast." Cancer Chemother Pharmacol **39**(4): 367-375.

Michaelis, C., R. Ciosk and K. Nasmyth (1997). "Cohesins: chromosomal proteins that prevent premature separation of sister chromatids." Cell **91**(1): 35-45.

Minchell, N. E., A. Keszthelyi and J. Baxter (2020). "Cohesin Causes Replicative DNA Damage by Trapping DNA Topological Stress." Mol Cell **78**(4): 739-751 e738.

Mizuguchi, T., G. Fudenberg, S. Mehta, J. M. Belton, N. Taneja, H. D. Folco, P. FitzGerald, J. Dekker, L. Mirny, J. Barrowman and S. I. S. Grewal (2014). "Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe." Nature **516**(7531): 432-435.

Mondal, N. and J. D. Parvin (2001). "DNA topoisomerase IIalpha is required for RNA polymerase II transcription on chromatin templates." Nature **413**(6854): 435-438.

Murayama, Y. and F. Uhlmann (2015). "DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism." Cell **163**(7): 1628-1640.

Murchie, A. I. and D. M. Lilley (1987). "The mechanism of cruciform formation in supercoiled DNA: initial opening of central basepairs in salt-dependent extrusion." Nucleic Acids Res **15**(23): 9641-9654.

Murugesapillai, D., M. J. McCauley, R. Huo, M. H. Nelson Holte, A. Stepanyants, L. J. Maher, 3rd, N. E. Israeloff and M. C. Williams (2014). "DNA bridging and looping by HMO1 provides a mechanism for stabilizing nucleosome-free chromatin." Nucleic Acids Res **42**(14): 8996-9004.

Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein and M. Snyder (2008). "The transcriptional landscape of the yeast genome defined by RNA sequencing." Science **320**(5881): 1344-1349.

Naughton, C., N. Avlonitis, S. Corless, J. G. Prendergast, I. K. Mati, P. P. Eijk, S. L. Cockroft, M. Bradley, B. Ylstra and N. Gilbert (2013). "Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures." Nat Struct Mol Biol **20**(3): 387-395.

Newlon, C. S. and J. F. Theis (1993). "The structure and function of yeast ARS elements." Curr Opin Genet Dev **3**(5): 752-758.

Nickol, J. and R. G. Martin (1983). "DNA stem-loop structures bind poorly to histone octamer cores." Proc Natl Acad Sci U S A **80**(15): 4669-4673.

Nobile, C., J. Nickol and R. G. Martin (1986). "Nucleosome phasing on a DNA fragment from the replication origin of simian virus 40 and rephasing upon cruciform formation of the DNA." Mol Cell Biol **6**(8): 2916-2922.

Nora, E. P., L. Caccianini, G. Fudenberg, K. So, V. Kameswaran, A. Nagle, A. Uebersohn, B. Hajj, A. L. Saux, A. Coulon, L. A. Mirny, K. S. Pollard, M. Dahan and B. G. Bruneau (2020). "Molecular basis of CTCF binding polarity in genome folding." Nat Commun **11**(1): 5612.

Nuebler, J., G. Fudenberg, M. Imakaev, N. Abdennur and L. A. Mirny (2018). "Chromatin organization by an interplay of loop extrusion and compartmental segregation." Proc Natl Acad Sci U S A **115**(29): E6697-E6706.

O'Sullivan, J. M., S. M. Tan-Wong, A. Morillon, B. Lee, J. Coles, J. Mellor and N. J. Proudfoot (2004). "Gene loops juxtapose promoters and terminators in yeast." Nat Genet **36**(9): 1014-1018.

Ocampo-Hafalla, M., S. Munoz, C. P. Samora and F. Uhlmann (2016). "Evidence for cohesin sliding along budding yeast chromosomes." Open Biol **6**(6).

Panday, A. and A. Grove (2017). "Yeast HMO1: Linker Histone Reinvented." Microbiol Mol Biol Rev **81**(1).

Patterton, H. G. and C. von Holt (1993). "Negative supercoiling and nucleosome cores. I. The effect of negative supercoiling on the efficiency of nucleosome core formation in vitro." J Mol Biol **229**(3): 623-636.

Patterton, H. G. and C. von Holt (1993). "Negative supercoiling and nucleosome cores. II. The effect of negative supercoiling on the positioning of nucleosome cores in vitro." J Mol Biol **229**(3): 637-655.

Pedersen, J. M., J. Fredsoe, M. Roedgaard, L. Andreasen, K. Mundbjerg, M. Kruhoffer, M. Brinch, M. H. Schierup, L. Bjergbaek and A. H. Andersen (2012). "DNA Topoisomerases maintain promoters in a state competent for transcriptional activation in Saccharomyces cerevisiae." PLoS Genet **8**(12): e1003128.

Postow, L., N. J. Crisona, B. J. Peter, C. D. Hardy and N. R. Cozzarelli (2001). "Topological challenges to DNA replication: conformations at the fork." Proc Natl Acad Sci U S A **98**(15): 8219-8226.

Pouokam, M., B. Cruz, S. Burgess, M. R. Segal, M. Vazquez and J. Arsuaga (2019). "The Rabl configuration limits topological entanglement of chromosomes in budding yeast." Sci Rep **9**(1): 6795.

Prunell, A. (1998). "A topological approach to nucleosome structure and dynamics: the linking number paradox and other issues." Biophys J **74**(5): 2531-2544.

Racko, D., F. Benedetti, J. Dorier and A. Stasiak (2018). "Transcription-induced supercoiling as the driving force of chromatin loop extrusion during formation of TADs in interphase chromosomes." Nucleic Acids Res **46**(4): 1648-1660.

Revyakin, A., C. Y. Liu, R. H. Ebright and T. R. Strick (2006). "Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching." Science **314**(5802): 1139-1143.

Rhee, K. Y., M. Opel, E. Ito, S. Hung, S. M. Arfin and G. W. Hatfield (1999). "Transcriptional coupling between the divergent promoters of a prototypic LysR-type regulatory system, the ilvYC operon of Escherichia coli." Proc Natl Acad Sci U S A **96**(25): 14294-14299.

Roedgaard, M., J. Fredsoe, J. M. Pedersen, L. Bjergbaek and A. H. Andersen (2015). "DNA Topoisomerases Are Required for Preinitiation Complex Assembly during GAL Gene Activation." PLoS One **10**(7): e0132739.

Rolef Ben-Shahar, T., S. Heeger, C. Lehane, P. East, H. Flynn, M. Skehel and F. Uhlmann (2008). "Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion." Science **321**(5888): 563-566.

Rovinskiy, N., A. A. Agbleke, O. Chesnokova, Z. Pang and N. P. Higgins (2012). "Rates of gyrase supercoiling and transcription elongation control supercoil density in a bacterial chromosome." PLoS Genet **8**(8): e1002845.

Saavedra, R. A. and J. A. Huberman (1986). "Both DNA topoisomerases I and II relax 2 micron plasmid DNA in living yeast cells." Cell **45**(1): 65-70.

Saka, K., A. Takahashi, M. Sasaki and T. Kobayashi (2016). "More than 10% of yeast genes are related to genome stability and influence cellular senescence via rDNA maintenance." Nucleic Acids Res **44**(9): 4211-4221.

Salceda, J., X. Fernandez and J. Roca (2006). "Topoisomerase II, not topoisomerase I, is the proficient relaxase of nucleosomal DNA." EMBO J **25**(11): 2575-2583.

Schneider, R. and R. Grosschedl (2007). "Dynamics and interplay of nuclear architecture, genome organization, and gene expression." Genes Dev **21**(23): 3027-3043.

Schober, H., V. Kalck, M. A. Vega-Palas, G. Van Houwe, D. Sage, M. Unser, M. R. Gartenberg and S. M. Gasser (2008). "Controlled exchange of chromosomal arms reveals principles driving telomere interactions in yeast." Genome Res **18**(2): 261-271.

Schultz, M. C., S. J. Brill, Q. Ju, R. Sternglanz and R. H. Reeder (1992). "Topoisomerases and yeast rRNA transcription: negative supercoiling stimulates initiation and topoisomerase activity is required for elongation." Genes Dev **6**(7): 1332-1341.

Seila, A. C., J. M. Calabrese, S. S. Levine, G. W. Yeo, P. B. Rahl, R. A. Flynn, R. A. Young and P. A. Sharp (2008). "Divergent transcription from active promoters." Science **322**(5909): 1849-1851.

Shlyakhtenko, L. S., V. N. Potaman, R. R. Sinden and Y. L. Lyubchenko (1998). "Structure and dynamics of supercoil-stabilized DNA cruciforms." J Mol Biol **280**(1): 61-72.

Simonis, M., P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. de Wit, B. van Steensel and W. de Laat (2006). "Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C)." Nat Genet **38**(11): 1348-1354.

Sinden, R. R., J. O. Carlson and D. E. Pettijohn (1980). "Torsional tension in the DNA double helix measured with trimethylpsoralen in living E. coli cells: analogous measurements in insect and human cells." Cell **21**(3): 773-783.

Sperling, A. S., K. S. Jeong, T. Kitada and M. Grunstein (2011). "Topoisomerase II binds nucleosome-free DNA and acts redundantly with topoisomerase I to enhance recruitment of RNA Pol II in budding yeast." Proc Natl Acad Sci U S A **108**(31): 12693-12698.

Stewart, A. F., R. E. Herrera and A. Nordheim (1990). "Rapid induction of c-fos transcription reveals quantitative linkage of RNA polymerase II and DNA topoisomerase I enzyme activities." Cell **60**(1): 141-149.

Stros, M., A. Bacikova, E. Polanska, J. Stokrova and F. Strauss (2007). "HMGB1 interacts with human topoisomerase IIalpha and stimulates its catalytic activity." Nucleic Acids Res **35**(15): 5001-5013.

Szilard, R. K., P. E. Jacques, L. Laramee, B. Cheng, S. Galicia, A. R. Bataille, M. Yeung, M. Mendez, M. Bergeron, F. Robert and D. Durocher (2010). "Systematic identification of fragile sites via genome-wide location analysis of gamma-H2AX." Nat Struct Mol Biol **17**(3): 299-305.

Tabuchi, H. and S. Hirose (1988). "DNA supercoiling facilitates formation of the transcription initiation complex on the fibroin gene promoter." J Biol Chem **263**(30): 15282-15287.

Takata, H., Y. Tanaka and A. Matsuura (2005). "Late S phase-specific recruitment of Mre11 complex triggers hierarchical assembly of telomere replication proteins in Saccharomyces cerevisiae." Mol Cell **17**(4): 573-583.

Tan-Wong, S. M., J. B. Zaugg, J. Camblong, Z. Xu, D. W. Zhang, H. E. Mischo, A. Z. Ansari, N. M. Luscombe, L. M. Steinmetz and N. J. Proudfoot (2012). "Gene loops enhance transcriptional directionality." Science **338**(6107): 671-675.

Thompson, M., R. A. Haeusler, P. D. Good and D. R. Engelke (2003). "Nucleolar clustering of dispersed tRNA genes." Science **302**(5649): 1399-1401.

Tolhuis, B., R. J. Palstra, E. Splinter, F. Grosveld and W. de Laat (2002). "Looping and interaction between hypersensitive sites in the active beta-globin locus." Mol Cell **10**(6): 1453-1465.

Trigueros, S. and J. Roca (2002). "Failure to relax negative supercoiling of DNA is a primary cause of mitotic hyper-recombination in topoisomerase-deficient yeast cells." J Biol Chem **277**(40): 37207-37211.

Tsochatzidou, M., M. Malliarou, N. Papanikolaou, J. Roca and C. Nikolaou (2017). "Genome urbanization: clusters of topologically co-regulated genes delineate functional compartments in the genome of Saccharomyces cerevisiae." Nucleic Acids Res **45**(10): 5818-5828.

Uhlmann, F., F. Lottspeich and K. Nasmyth (1999). "Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1." Nature **400**(6739): 37-42.

Uuskula-Reimand, L., H. Hou, P. Samavarchi-Tehrani, M. V. Rudan, M. Liang, A. Medina-Rivera, H. Mohammed, D. Schmidt, P. Schwalie, E. J. Young, J. Reimand, S. Hadjur, A. C. Gingras and M. D. Wilson (2016). "Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders." Genome Biol **17**(1): 182.

Wang, J. C. (2002). "Cellular roles of DNA topoisomerases: a molecular perspective." Nat Rev Mol Cell Biol **3**(6): 430-440.

Wang, Y., S. Maharana, M. D. Wang and G. V. Shivashankar (2014). "Super-resolution microscopy reveals decondensed chromatin structure at transcription sites." Sci Rep **4**: 4477.

Warren, A. C. and P. R. Cook (1978). "Supercoiling of DNA and nuclear conformation during the cell-cycle." J Cell Sci **30**: 211-226.

Wellinger, R. J. and V. A. Zakian (2012). "Everything you ever wanted to know about Saccharomyces cerevisiae telomeres: beginning to end." Genetics **191**(4): 1073-1105.

Xu, Z. Y., W. Wei, J. Gagneur, F. Perocchi, S. Clauder-Munster, J. Camblong, E. Guffanti, F. Stutz, W. Huber and L. M. Steinmetz (2009). "Bidirectional promoters generate pervasive transcription in yeast." Nature **457**(7232): 1033-U1037.

Zechiedrich, E. L. and N. Osheroff (1990). "Eukaryotic topoisomerases recognize nucleic acid topology by preferentially interacting with DNA crossovers." EMBO J **9**(13): 4555-4562.

Zhang, N., S. G. Kuznetsov, S. K. Sharan, K. Li, P. H. Rao and D. Pati (2008). "A handcuff model for the cohesin complex." J Cell Biol **183**(6): 1019-1031.

Appendix

Achar, Y. J., M. Adhil, R. Choudhary, N. Gilbert and M. Foiani (2020). "Negative supercoil at gene boundaries modulates gene topology." Nature **577**(7792): 701-705.

Author contributions: Achar, Y. J and M. Foiani designed the experiments, interpreted results and prepared the manuscript. Achar, Y. J and M. Adhil performed the experiments. M. Adhil Performed statistical and computational analysis. R. Choudhary provided technical input and N. Gilbert provided bTMP and technical input for supercoil analysis.

Article

Negative supercoil at gene boundaries modulates gene topology

https://doi.org/10.1038/s41586-020-1934-4

Received: 30 April 2019

Accepted: 25 November 2019

Published online: 22 January 2020

Yathish Jagadheesh Achar[1]*, Mohamood Adhil[1], Ramveer Choudhary[1], Nick Gilbert[2] & Marco Foiani[1,3]*

Transcription challenges the integrity of replicating chromosomes by generating topological stress and conflicts with forks[1,2]. The DNA topoisomerases Top1 and Top2 and the HMGB family protein Hmo1 assist DNA replication and transcription[3–6]. Here we describe the topological architecture of genes in *Saccharomyces cerevisiae* during the G1 and S phases of the cell cycle. We found under-wound DNA at gene boundaries and over-wound DNA within coding regions. This arrangement does not depend on Pol II or S phase. Top2 and Hmo1 preserve negative supercoil at gene boundaries, while Top1 acts at coding regions. Transcription generates RNA–DNA hybrids within coding regions, independently of fork orientation. During S phase, Hmo1 protects under-wound DNA from Top2, while Top2 confines Pol II and Top1 at coding units, counteracting transcription leakage and aberrant hybrids at gene boundaries. Negative supercoil at gene boundaries prevents supercoil diffusion and nucleosome repositioning at coding regions. DNA looping occurs at Top2 clusters. We propose that Hmo1 locks gene boundaries in a cruciform conformation and, with Top2, modulates the architecture of genes that retain the memory of the topological arrangements even when transcription is repressed.

RNA polymerases generate positive and negative supercoils ahead and behind transcription bubbles, respectively[7]. Positive supercoiling accumulates in front of replication forks and precatenanes are generated behind forks[8,9]. TMP (4,5′,8-trimethylpsoralen) has been used to map DNA supercoiling[10–13], as intercalation of psoralen is proportional to negative superhelical tension[14]. Using biotinylated TMP (bTMP)[12], we investigated the topology of transcribed genes and the contributions of Top1, Top2 and Hmo1 to maintenance of the topological architecture of transcription units.

Topological context of Pol II genes

We analysed the distributions of Rpb3 (a Pol II subunit), Top2, Top1 and Hmo1 in S phase and performed a meta-analysis on Pol II-transcribed genes (Fig. 1a). Rpb3 accumulated at open reading frames (ORFs), peaking at transcription start sites (TSSs) and transcription termination sites (TTSs); this probably reflects slow transcription modes where transcription begins, and where transcription-coupled transactions occur at termination[15]. Top2 and Hmo1 accumulated upstream and downstream of ORFs. Top1 was confined within the ORFs, accumulating close to TTSs.

Using bTMP, we mapped negative and positive supercoil and stable regions[12] (Extended Data Fig. 1a, b). The three topological clusters were distributed near-equally (Extended Data Fig. 1c). Negative supercoiled regions were mainly found at intergenic regions (49%), whereas transcribed units exhibited a positive supercoiled context (40%; Fig. 1b, c). Nucleosome-occupied regions were distributed near-equally within the three topological territories (Extended Data Fig. 1c). Negative supercoil mirrored Top2 clusters (Fisher's exact test $P < 1 \times 10^{-5}$) and positive supercoil reflected Top1 distribution (Fisher's exact test $P < 1 \times 10^{-20}$). Gene topology was comparable in cells during G1 and S phase (Fig. 1b, c, Extended Data Fig. 1d, e).

We analysed the topological profiles of conditionally expressed genes such as *ASF2* (transcribed in S phase) and the galactose-inducible gene cluster (Extended Data Fig. 1f, g). We found comparable bTMP profiles at *ASF2* in G1 (repressing conditions) and S, and at the Gal genes in cells cultured with glucose (repressing conditions) and galactose. We analysed the locus containing the highly expressed *LEU2* gene and the two moderately expressed *NFS1* and *DCC1* genes (Extended Data Fig. 1h). Rpb3 accumulated at *LEU2* but was undetectable at *NFS1* and *DCC1*. However, Top1 accumulated at *LEU2*, as well as at *NFS1* and *DCC1*. Top2 and Hmo1 were present at gene boundaries. Hence, the topological context of Pol II-transcribed genes does not depend on Pol II and the negative supercoil context at gene boundaries does not depend on Top2, as Top2 is recruited after G1 phase[16].

We compared supercoil distribution among genes that showed high, medium or low expression (Extended Data Fig. 2a). Highly expressed genes accumulated more Top2 and negative supercoil, compared to the other two classes (Fig. 1d, e). Conversely, highly expressed genes exhibited less positive supercoil (Extended Data Fig. 2b). Accumulation of Pol II and Top1 mirrored the levels of expression (Fig. 1d), whereas distribution of Hmo1 was comparable in genes of all three levels of expression (Fig. 1d). Hence, distribution of under-wound DNA at Pol II gene boundaries is enhanced in highly expressed genes.

[1]IFOM (Fondazione Istituto FIRC di Oncologia Molecolare), Milan, Italy. [2]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [3]Università degli Studi di Milano, Milan, Italy. *e-mail: yathish.achar@ifom.eu; marco.foiani@ifom.eu

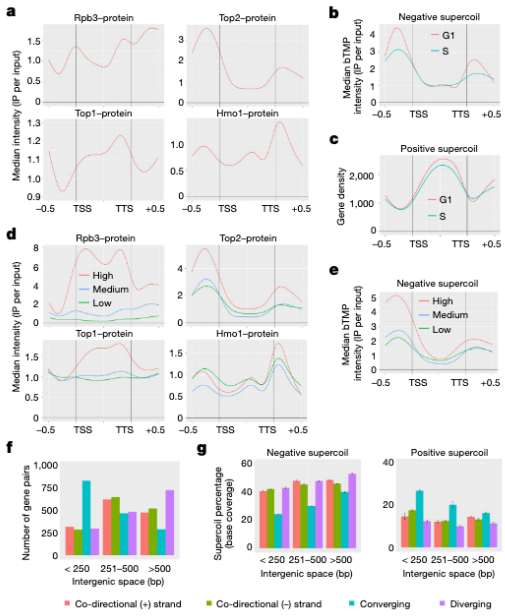


Fig. 1 | Topological context of Pol II genes. Chromatin immunoprecipitation (ChIP)-on-chip was carried out in cells released from G1 into S phase. Pol II-coding regions (replicates $n = 2$; meta-gene analysis $n = 6,706$ genes) were scaled to 1 kb and the flanking 0.5 kb from TSSs and TTSs were plotted against median intensity on the *y*-axis. **a**, Meta-gene plot showing accumulation of Pol II (Rpb3–10× Flag), Top2 (Top2–10× Flag), Top1 (Top1–10× Flag) and Hmo1 (Hmo1–10× Flag). **b**, Meta-gene plot for negative supercoil in G1 and S, plotted against median bTMP intensity. **c**, Positive supercoil distribution in Pol II genes plotted against average gene density on the *y*-axis. **d**, Pol II genes were grouped into three categories; high, medium and low expression based on the fragments per million kilobases (FPKM) value from RNA sequencing carried out in S phase at 28 °C. Meta-gene plots for three categories of gene expression for Pol II, Top2, Top1 and Hmo1. **e**, Negative supercoil distribution in high-, medium- and low-expression genes. **f**, Pol II genes were grouped according to their orientation with respect to neighbouring genes as: co-directional (+ strand; $n = 1,453$ gene pairs), co-directional (− strand; $n = 1,415$ gene pairs), converging ($n = 1,590$ gene pairs) and diverging ($n = 1,512$ gene pairs). The numbers of gene pairs at different intergenic spaces (<250 bp = 1,729 gene pairs, 251–500 bp = 2,224 gene pairs and >500 bp = 2,010 gene pairs) were plotted with respect to their orientation. **g**, Base coverage percentage of supercoil accumulation at different intergenic spaces with respect to gene pairs grouped according to orientation (two replicates, mean ± s.d.).

We compared the supercoil context at intergenic spaces with respect to gene orientation by grouping Pol II genes into co-directional (plus and minus strands), converging and diverging classes (Fig. 1f). Intergenic spaces between converging genes were smaller than in the other directional classes. Diverging genes exhibited larger intergenic spaces. Converging genes accumulated more positive supercoil at intergenic spaces, at the expense of negative supercoil (Fig. 1g). Accordingly, converging intergenic regions exhibited lower Top2 binding, whereas Top1 binding was not affected (Extended Data Fig. 2c). Hence, convergent and divergent transcription have imposed specific topological contexts at intergenic spaces.

Top2 and Hmo1 contribute to gene topology

Temperature sensitive *top2-1* mutants exhibited a reduction in negative supercoil at gene boundaries and an increase in positive supercoil at the same regions (Fig. 2a, Extended Data Fig. 2d, h). In G1 phase, *top2-1* mutants did not show a reduction in negative supercoil (Extended Data Fig. 2e). *top1Δ* cells did not show changes in the topological context of gene boundaries or transcribed regions (Fig. 2a, Extended Data Fig. 2f, h). Like *top2-1* mutants, *top1Δtop2-1* mutants exhibited a decrease in negative supercoil at gene boundaries (Fig. 2a), although the accumulation of positive supercoil at the same regions was lower than in *top2-1* cells (Extended Data Fig. 2g); *top1Δtop2-1* mutants accumulated less positive supercoil at transcribed regions than did wild-type cells, suggesting that, during transcription, the two topoisomerases can substitute for each other in maintaining a positive supercoiled context. Localization of Top1 in wild-type cells was restricted to coding regions, but in *top2-1* mutants it accumulated at gene boundaries (Fig. 2b, Extended Data Fig. 2i). Thus, Top2 restricts Top1 at transcribed regions and *top2* mutants exhibit the unscheduled relocation of Top1 at gene boundaries, which can account for the local increase in positive supercoil. Moreover, in *top2-1* mutants, Pol II accumulated more at gene boundaries than in wild-type cells (Fig. 2c, Extended Data Fig. 2i). Hence, Top2 confines the transcription apparatus within the coding regions.

hmo1Δ cells exhibited a reduction in negative supercoil and accumulation of positive supercoil at gene boundaries, resembling *top2-1* mutants (Fig. 2d, Extended Data Fig. 3a, d). Ablation of Hmo1 in *top2-1* mutants restored a wild-type-like topological context at transcribed genes (Fig. 2d, Extended Data Fig. 3b, d). The distribution of Top1 in *hmo1Δ* and *hmo1Δtop2-1* mutants was similar to that in wild-type cells (Extended Data Fig. 3c, e). Thus, the gene topological profiles of *hmo1Δ* and *top2-1* are comparable, but *hmo1Δ*, unlike *top2-1* mutation, does not cause accumulation of Top1 at gene boundaries. It is possible that, in *top2-1* mutants, accumulation of Top1 at gene boundaries depends on DNA substrates generated by Hmo1.

Top2 restricts RNA–DNA hybrids within ORFs

Using the SF9 antibody[17,18], we investigated whether accumulation of RNA–DNA hybrids reflected a specific topological context. In wild-type cells, hybrids were distributed within ORFs and peaked at TTSs (Extended Data Fig. 4a). Their accumulation did not correlate with gene expression levels (Extended Data Fig. 4b). RNaseH and the Rrm3 and Sen1 helicases counteract hybrid accumulation[19]. In *rnh1Δ* mutant cells, hybrids accumulated throughout Pol II gene units, whereas in *rrm3Δ* and *sen1*^{cl} mutants hybrids accumulated at TTS sites (Extended Data Fig. 4c); this is consistent with the function of Rrm3 in dismantling RNA transcripts while travelling on the lagging strand[20] and with the role of Sen1 in facilitating transcription termination[21]. These observations suggest that RNA–DNA hybrids represent a physiological intermediate during transcription and are confined within coding regions, and that their accumulation close to TTSs may reflect the slow-down of Pol II elongation at termination[22].

In *top2* mutants, specifically in S phase, hybrids accumulated at gene boundaries, where there is a reduction in negative supercoil (Fig. 2e). *top1top2* double mutants resembled *top2* mutants (Extended Data Fig. 4e). *top1Δ* cells accumulated hybrids throughout the gene bodies (Fig. 2e), perhaps owing to frequent Pol II pausing and back-tracking; because the viability of *top1Δ* cells depends on Top2, it is possible that *top1Δ* cells phenocopy a Top2 defect, leading to Pol II leakage and accumulation of hybrids at gene boundaries. Previous findings implicated Top1 in preventing hybrid accumulation[23]. Hence, Top2 counteracts hybrid accumulation and, in *top2* mutants, the accumulation of hybrids at flanking regions reflects the local decrease in negative supercoil and aberrant Pol II transcription. *hmo1Δ* cells exhibited a marked reduction in hybrid accumulation compared to wild-type cells, and *hmo1Δtop2-1*

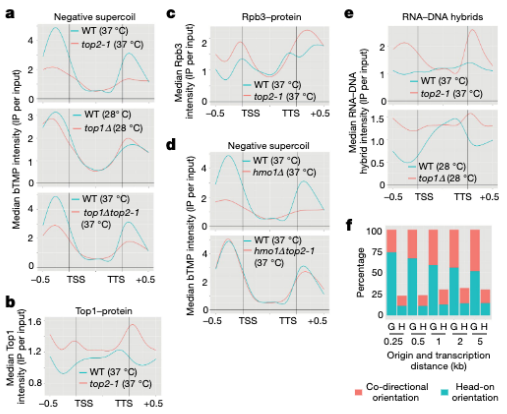


Fig. 2 | Top2 and Hmo1 contribute to gene topology. G1 cells were released at 28 °C into S phase and temperature shifted to 37 °C for *top2-1* mutants (replicates $n = 2$; meta-gene analysis $n = 6{,}706$ genes). **a**, Meta-gene profiles for negative supercoil comparison in wild-type (WT), *top2-1*, *top1Δ* and *top1Δtop2-1* cells. **b**, **c**, Meta-gene profiles comparing wild-type and *top2-1* cells for accumulation of Top1 protein (Top1–10× Flag) (**b**) or Pol II (Rpb3–10× Flag) (**c**). **d**, Meta-gene profile comparison for negative supercoil in wild-type, *hmo1Δ* and *hmo1Δtop2-1* cells. **e**, Meta-gene profiles for RNA–DNA hybrid comparison in wild-type, *top2-1*and *top1Δ* cells. **f**, Percentage of genes (G) and RNA–DNA hybrid (H) in either head-on or co-directional orientation with respect to replication forks. Genes were grouped with respect to their distance (0.25 kb, $n = 140$ genes; 0.5 kb, $n = 235$ genes; 1 kb, $n = 347$ genes; 2 kb, $n = 539$ genes; 5 kb, $n = 1{,}121$ genes) and direction (head-on or co-directional) from replication origin.

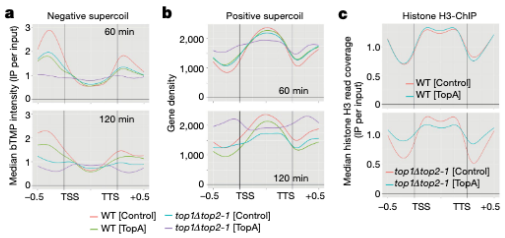


Fig. 3 | Negative supercoil disruption causes disarray in nucleosome occupancy. Wild-type and *top1Δtop2-1* mutants either harbouring control plasmid or expressing *E.coli* DNA TopA plasmid were grown at 28 °C and shifted to 37 °C for 60 or 120 min to inactivate Top2 (replicates $n = 2$; meta-gene analysis $n = 6{,}706$ genes). **a**, Meta-gene profiles for negative supercoil comparison in wild-type [Control plasmid], wild-type [TopA], *top1Δtop2-1* [Control] and *top1Δtop2-1* [TopA] at 60 min and 120 min at restrictive temperature. **b**, Meta-gene profile for positive supercoil accumulation. **c**, Meta-gene profiles of histone H3 in wild-type and *top1Δtop2-1* mutants with control or TopA plasmids plotted against median read coverage on the *y*-axis.

mutants behaved similarly to *hmo1Δ* cells (Extended Data Fig. 4f, g). Hence, ablation of *HMO1* also rescued the aberrant accumulation of hybrids at flanking regions in *top2* mutants.

We investigated whether a clash between forks and transcribed genes might influence the accumulation of hybrids, by analysing 173 efficient replication origins[24]. Transcription units in a head-on or co-directional orientation with replication forks within 0.25, 0.5, 1, 2 or 5 kb of the origin point were selected. There was a significant enrichment of transcribed genes oriented head-on with replication forks (Fig. 2f); this reflects the overlap between the signals that specify transcription termination and those that promote replication initiation[25]. However, the relative accumulation of hybrids in the head-on and the co-directional classes of genes were comparable (Fig. 2f, Extended Data Fig. 4h). Notably, the intergenic regions of converging genes were prone to accumulate hybrids, while this was not the case for the intergenic regions of divergent genes (Extended Data Fig. 4i).

Negative supercoil affects gene architecture

To validate the previous observations, we expressed *Escherichia coli* DNA topoisomerase I (TopA). TopA expression in *top1top2* mutants depletes negative supercoil in plasmids[26]. Wild-type and *top1Δtop2-1* cells harbouring either control vector or TopA-expressing plasmids were analysed after 60 and 120 min at the restrictive temperature for *top2-1* mutation (Fig. 3a, b). TopA expression in wild-type cells showed a reduction in negative supercoil at ORF-flanking regions and, in *top1Δtop2-1* cells, nearly abolished the negative supercoil at flanking regions (Fig. 3a, Extended Data Fig. 5a). Hence, the presence of Hmo1 at gene boundaries in *top1Δtop2-1* double mutants does not prevent TopA from resolving negative supercoil. TopA acts on negative supercoil to convert it into positive supercoil[26]. Accordingly, the disappearance of negative supercoil at flanking regions paralleled the progressive accumulation of over-wound DNA at the same location (Fig. 3b). Expression of TopA caused a reduction in positive supercoil at transcribed regions in wild-type cells, whereas it had the opposite effect in *top1Δtop2-1* mutants (Fig. 3b). This could result from the diffusion of supercoil waves across the entire gene bodies, perhaps owing to the destruction of topological or architectural confinements. Binding of Hmo1 was reduced in *top1Δtop2-1* mutants compared to wild-type cells and was nearly abolished in *top1Δtop2-1* cells expressing TopA (Extended Data Fig. 5b), indicating that association of Hmo1 with gene boundaries depends on negative supercoil.

In wild-type cells, histone H3 was distributed at transcribed units but was less abundant at gene boundaries (Fig. 3c). *top1top2* mutants resembled wild-type cells, suggesting that the aberrant topological context in the double mutants did not affect the nucleosome context. Expression of TopA did not alter nucleosome positioning and distribution in wild-type cells, but in *top1top2* mutants it caused reduction of H3 distribution (Extended Data Fig. 5c). Moreover, H3 redistributed as its levels increased at flanking regions and, concomitantly, decreased at transcribed units, starting from position +2 (Fig. 3c, Extended Data Fig. 5d). Hence, expression of TopA caused an increase in positive supercoil followed by diffusion of supercoil waves across the entire gene body and massive nucleosome repositioning.

Top2 mediates chromatin loop formation

Using the chromatin interaction analysis by paired-end tag sequencing (ChIA-PET)[27] method, we investigated whether Top2 mediates the formation of chromatin loops. We used Top2 as bait in S phase cells. Following DNA sequencing, we acquired one million independently mapped paired end tags (PETs) (Extended Data Fig. 6a, b) and, by keeping a 1-kb minimum distance, we obtained 1,887 inter-ligation PET clusters (Extended Data Fig. 6b, c). The lengths of the Top2-mediated loops varied; some were larger than 10 kb (~100 interactions), while the majority of loops were between 1,500 and 2,000 bp in size with a median of 1,900 bp (Fig. 4a, b, Extended Data Fig. 6d). Sixty-four per cent of the interactions corresponded to previously described Top2-binding sites[3] and, for the majority of the interactions (66%), Top2 was found only at one end of the loop (Extended Data Fig. 6e). Overall, 45% of the Pol II genes were located within loops. Several loops were organized

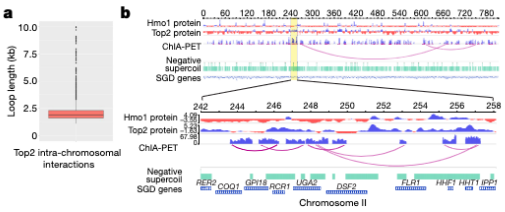


Fig. 4 | Top2 mediates chromatin loop formation. a, Box plot showing loop size distribution (n = 1,505 loops; min = 1.1 kb; max = 662 kb; median = 1.9 kb; 25th percentile = 1.6 kb; 75th percentile = 2.4 kb). **b**, Genome browser view of how Top2 mediates chromatin interactions on chromosome II, along with Top2 and Hmo1 protein chip data. Highlighted area (chr. II: 242000 to 258000) is enlarged below.

in clusters, with 31% of loops containing more than one gene and 51% of loops containing a single gene (Fig. 4b, Extended Data Fig. 6e, f).

Discussion

We have shown that Top1 localizes at coding regions. Top2 instead acts at negatively supercoiled gene boundaries and engages genes in loop-like structures, bringing promoters and terminators into proximity[28]. Multiple twin topological domains are likely to be generated within the gene loops by waves of Pol II complexes[29]. Whereas in prokaryotes negative supercoil behind the first transcription bubble might adsorb the positive supercoil generated by the next approaching Pol II complex[30], in eukaryotes, it may enable nucleosome assembly following Pol II passage[31]. Hence, eukaryotic RNA polymerase progression might depend strongly on Top1 in resolving topological stress in front of Pol II. Accordingly, accumulation of Top1 at coding regions depends on transcription levels. Coding regions exhibit a positive supercoiled context, even when transcription is repressed, implying that genes retain a 'memory' of a topological architecture that does not reflect the dynamics of elongating Pol II[7,32].

The negatively supercoiled regions that flank ORFs are refractory to nucleosome formation; TopA depletes negative supercoil specifically at these regions. Hence, gene boundaries exhibit an ideal topological context to 'breathe out' and undergo alternative structural transitions[33]. Nucleosome-free negatively supercoiled regions can form pseudo-cruciform structures[34], characterized by two B-DNA duplex arms and two intra-strand plectonemic arms in a non-B-DNA conformation (Extended Data Fig. 7a). Such structures can branch-migrate, thereby modulating the extension of the intra-strand plectonemic duplexes[35]. Like other HMG box proteins[36], Hmo1 binds four-way junctions with high affinity. Moreover, it stabilizes nucleosome-free regions, and dimerizes to promote DNA bridging[37]. We propose that Hmo1 locks cruciform DNA and thereby counteracts branch migration and nucleosome formation. Stable negatively supercoiled gene boundaries in a cruciform conformation might help to insulate the topological architecture of gene loops to facilitate elongation of the multiple Pol II complexes, allowing efficient recycling of Pol II from TTS to TSS. Dimerization of Hmo1[37] may promote gene looping, even in G1, without the mediation of Top2. Pol II movement and transcription-coupled processes, such as gene gating and/or splicing, might also contribute to gene looping by extruding portions of the transcribed DNA[38]. In S phase, Top2 would act at the loop base, probably to counteract the disruptive potential of incoming forks and/or to reset gene topology after fork passage (Extended Data Fig. 7b). Notably, Top2-dependent DNA loops can contain more than one transcription unit, and can be organized in clusters, thus generating complex topological structures.

S phase cells accumulate RNA–DNA hybrids at ORFs in about 45% of genes, independent of gene expression levels and of the direction of transcription–replication. However, the intergenic regions of converging genes exhibit a bias for hybrid accumulation. Pol II back-tracking during elongation, and a slow-down of Pol II during termination[39,40], could account for the formation of hybrids at ORFs under physiological conditions. The under-wound DNA behind Pol II can easily accommodate RNA–DNA pairing[41], and might even muffle the negative supercoil generated by Pol II movement. Our data suggest that hybrid formation is a physiological event, intrinsic to the topological dynamics generated by transcription and co-transcriptional processes. However, converging genes might generate the context for unscheduled genotoxic events, as in the case of CSR-activated B cells[42].

Our model (Extended Data Fig. 7b) leads to the following predictions. (i) Recruitment of Hmo1 at gene boundaries would depend on their negative supercoil state; Hmo1 would then generate stable negative supercoiled cruciforms at gene boundaries (Extended Data Fig. 7a). Hmo1 is always found at negative supercoiled and nucleosome-free regions flanking ORFs, whereas Top2 is recruited in S phase. Counteracting negative supercoil at gene boundaries prevents recruitment of Hmo1. Without Hmo1, cruciforms would be unstable but remain in a negative supercoiled state, becoming an ideal substrate for Top2 (Extended Data Fig. 7c). Accordingly, inactivation of Top2 in *hmo1* cells rescues negative supercoil at gene boundaries. (ii) In *top2* mutants, negative supercoil decreases at ORF-flanking regions, probably owing to the unscheduled and massive recruitment of Top1 at gene boundaries. Notably, non-B-DNA structures can be a substrate for Top1[43], and Top1 can efficiently relax both positive and negative supercoil[44]. Hence, Hmo1 cannot protect cruciforms from Top1 activity when *top2* is mutated, implying that, in *top2* mutants, Top1 might cause genotoxic events at Hmo1-locked cruciforms, such as extensive nicking and/or knotting[45] (Extended Data Fig. 7c). Notably, deletion of *HMO1* in *top2* mutants, besides alleviating *top2* temperature sensitivity[3], prevents relocation of Top1 at flanking regions, and *hmo1top2* double mutants exhibit a wild-type-like topological context. (iii) In S phase *top2* mutants, Pol II leaks outside the canonical transcribed regions. This aberrant Pol II distribution is likely to reflect the inability of *top2* mutants to recycle Pol II from TTSs to TSSs, owing to the loss of proximity between promoters and terminators. In this view, Top2 might protect the gene loop structure from incoming forks. (iv) The aberrant Pol II distribution in *top2* mutants may also account for hybrid accumulation at gene boundaries. In *top2* mutants, hybrid accumulation downstream of ORFs may result from aberrant transcription termination, while upstream of ORFs it might be facilitated by the Top1-mediated processing of cruciform DNA. In fact, *top2top1* mutants exhibit fewer hybrids than single *top2* mutants. Another possibility is that Top2 defects promote aberrant antisense transcription initiation events close to TSSs.

The hybrids that accumulate at gene boundaries in *top2* mutants may generate genotoxic events and the unscheduled synthesis of small RNA species, and might contribute to absorbance of the negative supercoil, thus implying that negative supercoil reduction at flanking regions may represent an indirect consequence of Top1 relocation.

Our observations suggest that Top1, Top2 and Hmo1 contribute to the topological architecture of transcribed genes, particularly in S phase when forks reset the topological states of chromosomes and their chromatin context. Interfering with the topological context of gene-flanking regions may cause a variety of pathological consequences, such as the generation of aberrant RNA species, the accumulation of RNA–DNA hybrids and alterations at the level of chromatin architecture.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-1934-4.

1. Bermejo, R., Lai, M. S. & Foiani, M. Preventing replication stress to maintain genome stability: resolving conflicts between replication and transcription. *Mol. Cell* **45**, 710–718 (2012).
2. García-Muse, T. & Aguilera, A. Transcription-replication conflicts: how they occur and how they are resolved. *Nat. Rev. Mol. Cell Biol.* **17**, 553–563 (2016).
3. Bermejo, R. et al. Genome-organizing factors Top2 and Hmo1 prevent chromosome fragility at sites of S phase transcription. *Cell* **138**, 870–884 (2009).
4. Sperling, A. S., Jeong, K. S., Kitada, T. & Grunstein, M. Topoisomerase II binds nucleosome-free DNA and acts redundantly with topoisomerase I to enhance recruitment of RNA Pol II in budding yeast. *Proc. Natl Acad. Sci. USA* **108**, 12693–12698 (2011).
5. Wang, J. C. Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell Biol.* **3**, 430–440 (2002).
6. Pedersen, J. M. et al. DNA topoisomerases maintain promoters in a state competent for transcriptional activation in *Saccharomyces cerevisiae*. *PLoS Genetics* **8**, e1003128 (2012).
7. Liu, L. F. & Wang, J. C. Supercoiling of the DNA template during transcription. *Proc. Natl Acad. Sci. USA* **84**, 7024–7027 (1987).
8. Postow, L., Crisona, N. J., Peter, B. J., Hardy, C. D. & Cozzarelli, N. R. Topological challenges to DNA replication: conformations at the fork. *Proc. Natl Acad. Sci. USA* **98**, 8219–8226 (2001).
9. Schvartzman, J. B. & Stasiak, A. A topological view of the replicon. *EMBO Rep.* **5**, 256–261 (2004).
10. Lal, A. et al. Genome scale patterns of supercoiling in a bacterial chromosome. *Nat. Commun.* **7**, 11055 (2016).
11. Bermúdez, I., García-Martínez, J., Pérez-Ortín, J. E. & Roca, J. A method for genome-wide analysis of DNA helical tension by means of psoralen-DNA photobinding. *Nucleic Acids Res.* **38**, e182 (2010).
12. Naughton, C. et al. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat. Struct. Mol. Biol.* **20**, 387–395 (2013).
13. Kouzine, F. et al. Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat. Struct. Mol. Biol.* **20**, 396–403 (2013).
14. Sinden, R. R., Carlson, J. O. & Pettijohn, D. E. Torsional tension in the DNA double helix measured with trimethylpsoralen in living *E. coli* cells: analogous measurements in insect and human cells. *Cell* **21**, 773–783 (1980).
15. Perales, R. & Bentley, D. "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol. Cell* **36**, 178–191 (2009).
16. Bermejo, R. et al. Top1- and Top2-mediated topological transitions at replication forks ensure fork progression and stability and prevent DNA damage checkpoint activation. *Genes Dev.* **21**, 1921–1936 (2007).
17. Boguslawski, S. J. et al. Characterization of monoclonal antibody to DNA. RNA and its application to immunodetection of hybrids. *J. Immunol. Methods* **89**, 123–130 (1986).
18. Chan, Y. A. et al. Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. *PLoS Genet.* **10**, e1004288 (2014).
19. Hamperl, S. & Cimprich, K. A. The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair* **19**, 84–94 (2014).
20. Rossi, S. E., Ajazi, A., Carotenuto, W., Foiani, M. & Giannattasio, M. Rad53-mediated regulation of Rrm3 and Pif1 DNA helicases contributes to prevention of aberrant fork transitions under replication stress. *Cell Rep.* **13**, 80–92 (2015).
21. Hazelbaker, D. Z., Marquardt, S., Wlotzka, W. & Buratowski, S. Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. *Mol. Cell* **49**, 55–66 (2013).
22. Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N. J. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* **516**, 436–439 (2014).
23. Tuduri, S. et al. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell Biol.* **11**, 1315–1324 (2009).
24. Fachinetti, D. et al. Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements. *Mol. Cell* **39**, 595–605 (2010).
25. Chen, S., Reger, R., Miller, C. & Hyman, L. E. Transcriptional terminators of RNA polymerase II are associated with yeast replication origins. *Nucleic Acids Res.* **24**, 2885–2893 (1996).
26. Gartenberg, M. R. & Wang, J. C. Positive supercoiling of DNA greatly diminishes mRNA synthesis in yeast. *Proc. Natl Acad. Sci. USA* **89**, 11461–11465 (1992).
27. Li, X. et al. Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions. *Nat. Protocols* **12**, 899–915 (2017).
28. O'Sullivan, J. M. et al. Gene loops juxtapose promoters and terminators in yeast. *Nat. Genet.* **36**, 1014–1018 (2004).
29. Levens, D., Baranello, L. & Kouzine, F. Controlling gene expression by DNA mechanics: emerging insights and challenges. *Biophys. Rev.* **8**, 259–268 (2016).
30. Rovinskiy, N., Agbleke, A. A., Chesnokova, O., Pang, Z. & Higgins, N. P. Rates of gyrase supercoiling and transcription elongation control supercoil density in a bacterial chromosome. *PLoS Genet.* **8**, e1002845 (2012).
31. Patterton, H. G. & von Holt, C. Negative supercoiling and nucleosome cores. I. The effect of negative supercoiling on the efficiency of nucleosome core formation *in vitro*. *J. Mol. Biol.* **229**, 623–636 (1993).
32. Ma, J., Bai, L. & Wang, M. D. Transcription under torsion. *Science* **340**, 1580–1583 (2013).
33. Kouzine, F. et al. Permanganate/S1 nuclease footprinting reveals non-B DNA structures with regulatory potential across a mammalian genome. *Cell Syst.* **4**, 344–356 (2017).
34. Lilley, D. M. DNA opens up—supercoiling and heavy breathing. *Trends Genet.* **4**, 111–114 (1988).
35. Murchie, A. I. & Lilley, D. M. Supercoiled DNA and cruciform structures. *Methods Enzymol.* **211**, 158–180 (1992).
36. Pöhler, J. R. G., Norman, D. G., Bramham, J., Bianchi, M. E. & Lilley, D. M. HMG box proteins bind to four-way DNA junctions in their open conformation. *EMBO J.* **17**, 817–826 (1998).
37. Murugesapillai, D. et al. DNA bridging and looping by HMO1 provides a mechanism for stabilizing nucleosome-free chromatin. *Nucleic Acids Res.* **42**, 8996–9004 (2014).
38. Tan-Wong, S. M., Wijayatilake, H. D. & Proudfoot, N. J. Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes Dev.* **23**, 2610–2624 (2009).
39. Saponaro, M. et al. RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell* **157**, 1037–1049 (2014).
40. Cheung, A. C. & Cramer, P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**, 249–253 (2011).
41. Drolet, M. et al. Overexpression of RNase H partially complements the growth defect of an *Escherichia coli* delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proc. Natl Acad. Sci. USA* **92**, 3526–3530 (1995).
42. Meng, F. L. et al. Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* **159**, 1538–1548 (2014).
43. Husain, A. et al. Chromatin remodeller SMARCA4 recruits topoisomerase 1 and suppresses transcription-associated genomic instability. *Nat. Commun.* **7**, 10549 (2016).
44. Fernández, X., Díaz-Ingelmo, O., Martínez-García, B. & Roca, J. Chromatin regulates DNA torsional energy via topoisomerase II-mediated relaxation of positive supercoils. *EMBO J.* **33**, 1492–1501 (2014).
45. Brown, P. O. & Cozzarelli, N. R. Catenation and knotting of duplex DNA by type 1 topoisomerases: a mechanistic parallel with type 2 topoisomerases. *Proc. Natl Acad. Sci. USA* **78**, 843–847 (1981).

Methods

Strains and growing conditions

All *S. cerevisiae* strains are W303 derivatives[46]. The relevant genotypes are shown in Extended Data Table 1. Strains were grown at 28 °C in YPD medium. G1 synchronization was carried out using 3–5 μg/ml of α factor. For S-phase samples, G1 cells were washed twice in YP medium and allowed to grow for 15 min in fresh medium. For temperature-sensitive strains, cells were allowed to grow for 10 min in fresh medium after G1 release, centrifuged and then dissolved in pre-warmed medium at 37 °C and allowed to grow for 15 min. Cell cycle progression into S phase was monitored by fluorescence-activated cell sorting (FACS) and budding profiles. For *E. coli* TopA expression, wild-type cells and *top1Δtop2-1* mutants harbouring either control or TopA expression plasmids were grown at 25 °C in synthetic medium lacking leucine. Cells were shifted to 37 °C for inactivation of Top2 after reaching 8×10^6 cells/ml concentration.

bTMP-ChIP

We adapted the previously described method to yeast[12]. Sodium azide (0.1%) was used to block cells and to ensure the preservation of the most prevalent topological context present at any given genomic position. We note that this method does not aim to study dynamic topological transitions. Permeabilized yeast cells were incubated with bTMP (400 μg per 2×10^9 cells) in the dark for 90 min and then cross-linked by 365 nm UV (800 mJ/cm^2) light to form adducts between two DNA strands. Cells were washed twice with ice-cold PBS and lysed in 1 ml of lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate) using Zirconia beads. The cross-linked chromatin was sheared to an average size of 500 bp by 6 × 15-s pulses using a Biorupter sonicator and DNA was purified. Purified DNA was incubated with Dynabeads MyOne streptavidin (Invitrogen cat. no. 65001) overnight at 4 °C. The beads were washed twice with each of the following buffers; wash buffer-I (20 mM Tris-HCL pH 8, 2 mM EDTA, 150 mM NaCl, 1% Triton X, 0.1% SDS), wash buffer-II (20 mM Tris-HCL pH 8, 2 mM EDTA, 500 mM NaCl, 1% Triton X, 0.1% SDS), wash buffer III (250 mM LiCl, 10 mM Tris pH 8.0, 0.5% Na-deoxycholate, 0.5% NP-40, 1 mM EDTA) and 1× TE (20 mM Tris pH 8.0, 2 mM EDTA). The bTMP–DNA complexes were eluted from the beads in 250 μl elution buffer (95% formamide, 10 mM EDTA) at 90 °C for 20 min and eluted samples were cleaned with a Qiagen PCR clean up kit. Input DNA was isolated from sheared chromatin input (1/100 of the material used for ChIP). For bTMP-ChIP with naked DNA, genomic DNA was isolated from Qiagen Genomic-tip 100/G (cat. no. 13343) and Genomic DNA Buffer Set (cat. no. 19060). Purified DNA was sheared to an average size of 500 bp by 6× 15-s pulses using a Biorupter sonicator. bTMP was added to purified DNA and incubated in the dark for 90 min and cross-linked with UV at 365 nm (800 mJ/cm^2). DNA was precipitated using isopropanol and washed with 70% ethanol. The dried pellet was dissolved in buffer (50 mM Tris pH 8.0, 10 mM EDTA, 0.1% SDS) and incubated with Dynabeads MyOne streptavidin (Invitrogen cat. no. 65001) overnight at 4 °C . Washing and elution was as described above. Both IP and input samples were processed as described in the in microarray section.

The procedure for bTMP titration is presented in Extended Data Fig. 8. bTMP binding normalization and the dispersion profile for bTMP are presented in Extended Data Fig. 9.

Protein ChIP

ChIP analysis for proteins was carried out as described[47] with few modifications. Cells were cross-linked with 1% formaldehyde in culture medium for 30 min at room temperature followed by quenching with 0.125 M glycine for 5 min. Cells were washed twice with ice-cold PBS and lysed in 1 ml of lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate) using Zirconia beads. Cross-linked chromatin was sheared to an average size of 500 bp by 6x 15-s pulses using a Biorupter sonicator. The lysate was then centrifuged to remove cell debris. The chromatin fraction was incubated with Dynabeads protein G beads (Invitrogen, cat. no. 10003D) coated with anti-Flag antibody (M2-antiflag, Sigma) overnight at 4 °C. The immune complexes were washed with the following buffers 2×; Chip-lysis buffer, high-salt lysis buffer (Chip-lysis buffer + 360 mM NaCl), Chip-wash buffer (250 mM LiCl, 10 mM Tris pH 8.0, 0.5% Na-deoxycholate, 0.5% NP-40, 1 mM EDTA) and 1× TE (20 mM Tris pH 8.0, 2 mM EDTA). The protein-DNA complexes were eluted from the beads in 250 μl elution buffer (1% SDS, 50 mM Tris pH 8.0, 10 mM EDTA) at 65 °C for 20 min followed by the addition of proteinase K to 500 μg/ml and overnight incubation at 65 °C. Input DNA was isolated from sheared chromatin input (1/100 of the material used for ChIP). Both IP and input samples were processed as mentioned in the section 'Microarray and data processing'.

DRIP-ChIP

DRIP-ChIP was performed using anti-DNA:RNA hybrid monoclonal mouse antibody S9.6 as previously described[18]. In brief, cells were cross-linked with 1% formaldehyde in culture medium for 20 min at room temperature followed by quenching with 0.125 M glycine for 5 min. Cells were washed twice with ice-cold PBS and lysed in 1 ml of lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate) using Zirconia beads. Cross-linked chromatin was sheared to an average size of 500 bp by 6× 15-s pulses using a Biorupter sonicator. The lysate was then centrifuged to remove cell debris. The chromatin fraction was incubated with Protein-A magnetic beads (Invitrogen, cat. no. 10001D) coated with anti-DNA:RNA hybrid S9.6 antibody[17] overnight at 4 °C . The immune complexes were washed with the following buffers 2×; Chip-lysis buffer, high-salt lysis buffer (Chip-lysis buffer + 360 mM NaCl), Chip-wash buffer (250 mM LiCl, 10 mM Tris pH 8.0, 0.5% Na-deoxycholate, 0.5% NP-40, 1 mM EDTA) and 1× TE (20 mM Tris pH 8.0, 2 mM EDTA). The RNA:DNA hybrid complexes were eluted from the beads in 250 μl elution buffer (1% SDS, 50 mM Tris pH 8.0, 10 mM EDTA) at 65 °C for 20 min followed by the addition of proteinase K to 500 μg/ml and overnight incubation at 65 °C. Input DNA was isolated from sheared chromatin input (1/100 of the material used for ChIP). Both IP and input samples were processed as mentioned in the section 'Microarray and data processing'.

Histone H3 ChIP sequencing

ChIP analysis for proteins was carried out as described previously[48]. In brief, cells were cross-linked with 1% formaldehyde in culture medium for 15 min at room temperature followed by quenching with 0.125 M glycine for 5 min. Cells were washed twice with ice-cold PBS and lysed in 1 ml of lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate) using Zirconia beads. Cross-linked chromatin was sheared to an average size of 200 bp in Covaris S220 Focused Ultrasonicators. The lysate was then centrifuged to remove cell debris. The chromatin fraction was incubated with Protein-G magnetic beads (Invitrogen, cat. no. 10003D) coated with anti-histone H3 antibody (Abcam, cat. no. ab1791) overnight at 4 °C . The immune complexes were washed twice with the following buffers; ChIP-lysis buffer, high-salt lysis buffer (ChIP-lysis buffer + 360 mM NaCl), ChIP-wash buffer (250 mM LiCl, 10 mM Tris pH 8.0, 0.5% Na-deoxycholate, 0.5% NP-40, 1 mM EDTA) and 1× TE (20 mM Tris pH 8.0, 2 mM EDTA). The protein–DNA complexes were eluted from the beads in 250 μl elution buffer (1% SDS, 50 mM Tris pH 8.0, 10 mM EDTA) at 65 °C for 20 min followed by the addition of proteinase K to 500 μg/ml and overnight incubation at 65 °C. Input DNA was isolated from sheared chromatin input (1/100 of the material used for ChIP).

For sequencing, IP and input ChIP sequencing (ChIP–seq) libraries were prepared according to the manufacturer's protocols for the Ion Proton sequencer (Thermo Fisher Scientific/Life Technologies). In brief, 10 ng of ChIP DNA was end repaired and adaptor ligated using the KAPA

Library Preparation Kit for Ion Torrent (KAPA Biosystems) and adaptor barcode Kapa Barcode Adaptors 9-24. After adaptor ligation, each sample was size selected using AMPure XP Bead (Beckman Coulter). An amplification reaction was set up in a final volume of 50 µl. A SPRI cleanup with a 1.5× bead:DNA ratio was performed after amplification and final libraries were eluted in 35 µl. Libraries were quantified on a Qubit fluorometer with HS DNA (Thermo Fisher Scientific/Life Technologies) and checked for size on an Agilent Bioanalyzer with an HS DNA kit (Agilent). Each size-selected library was diluted to a final concentration of 11 pM and clonally amplified using the Ion Proton Hi-QTemplate Kit (Thermo Fisher Scientific/Life Technologies) with IonOneTouch 2 instrument (Thermo Fisher Scientific/Life Technologies). After emulsion PCR, DNA-positive ion sphere particles (ISPs) were recovered and enriched according to standard protocols with the IonOneTouch ES Instrument (Thermo Fisher Scientific/Life Technologies). A sequencing primer was annealed to DNA-positive ISPs and the sequencing polymerase bound, before loading of ISPs into Ion P1 sequencing chips. Sequencing of the samples was conducted according to the Ion Proton Hi-Q Sequencing Kit protocol. One P1 sequencing chip with six libraries was loaded and run on an Ion Proton sequencer.

RNA sequencing

Total RNA was isolated from 5×10^7 cells with the RNeasy Mini Kit (50) (Qiagen cat. no. 74104). Prior to library preparation, cytoplasmic and mitochondrial ribosomal RNA was removed using the Ribo-Zero Gold rRNA Removal Kit (Yeast) (Illumina, cat. no. MRZY1324). Libraries for RNA sequencing were prepared according to the manufacturer's protocols for transcriptome sequencing with the Ion Proton sequencer (Thermo Fisher Scientific/Life Technologies). In brief, 1 µg total RNA was poly-A-selected using the Dynabeads mRNA Direct Micro Purification kit (Thermo Fisher Scientific, cat. no. 61021) according to the manufacturer's protocol. About 50 ng of poly-A RNA was used to prepare strand-specific barcoded RNA libraries with the Ion Total RNA-Seq kit v.2.0 (Thermo Fisher Scientific, cat. no. 4475936). In brief, poly-A RNA was fragmented with RNase III and purified with Nucleic Acid Binding Beads. After purification, the poly-A RNA fragments were hybridized and ligated with Ion Adaptor and subsequently reverse transcribed for cDNA preparation. cDNAs were amplified with Ion Torrent barcoded primers and purified with Nucleic Acid Binding Beads. Final libraries were quantified on a Qubit fluorometer with HS DNA (Thermo Fisher Scientific) and checked for size on an Agilent Bioanalyzer with an HS DNA kit (Agilent). Four barcoded libraries were pooled together on an equimolar basis at a final concentration of 11 pM and clonally amplified using the Ion Proton Hi-QTemplate Kit (Thermo Fisher Scientific, cat no. A26434) with IonOneTouch 2 instrument (Thermo Fisher Scientific/Life Technologies). After emulsion PCR, DNA-positive ISPs were recovered and enriched by standard protocols with the IonOneTouch ES Instrument (Thermo Fisher Scientific/Life Technologies). A sequencing primer was annealed to DNA-positive ISPs and the sequencing polymerase bound, before loading of ISPs into Ion P1 sequencing chips. Sequencing of the samples was conducted according to the Ion Proton Hi-Q Sequencing Kit (Thermo Fisher Scientific, cat. no. A26433) Protocol on Ion Proton instrument.

ChIA-PET

We adopted the previously described method[27]. Cells were cross-linked with 1% formaldehyde in culture medium for 30 min at room temperature followed by quenching with 0.125 M glycine for 5 min. Cells were washed twice with ice-cold PBS and lysed in 1 ml of lysis buffer (50 mM HEPES- KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate) using Zirconia beads. Cross-linked chromatin was sheared to an average size of 500 bp by 6×15-s pulses using a Biorupter sonicator. The lysate was then centrifuged to remove cell debris. The chromatin fraction was incubated with Dynabeads protein G beads (Invitrogen, cat no 10003D) coated with anti-Flag antibody (M2-antiflag, Sigma) overnight at 4 °C . The immune complexes were washed twice with the following buffers: ChIP-lysis buffer, high-salt lysis buffer (ChIP-lysis buffer + 360 mM NaCl), ChIP-wash buffer (250 mM LiCl, 10 mM Tris pH 8.0, 0.5% Na-deoxycholate, 0.5% NP-40, 1 mM EDTA) and 1× TE (20 mM Tris pH8.0, 2 mM EDTA). Beads were pooled and end-repair was carried out using T4 DNA polymerase (NEB, cat. no. M0203L) by rotating on a Intelli-Mixer at 37 °C for 40 min. Beads were washed 3× with ice-cold ChIA-PET wash buffer (10 mM Tris pH 7.4, 1 mM EDTA, 500 mM NaCl). A-tailing was carried out using Klenow fragment (3′-5′ exo-) (NEB, cat. no. M0212M) in the presence of 100 uM dATP by rotating on a Intelli-Mixer at 37 °C for 50 min. Beads were washed 3× with ice cold ChIA-PET wash buffer (10 mM Tris pH 7.4, 1 mM EDTA, 500 mM NaCl). For proximity ligation, a bridge linker was prepared by annealing Linker-F and Linker-R (HPLC purified (250 nmole) from IDT (Integrated DNA Technologies): bridge linker-F: 5′- /5Phos/CGCGATATC/iBIOdT/TATCTGACT -3′; bridge linker-R: 5′- /5Phos/GTCAGATAAGATATCGCGT -3′.

Proximity ligation was carried out using T4 DNA ligase (NEB cat. no. M0202M), in the presence of bridge linker at a concentration of 0.57 ng/µl by rotating at 16 °C overnight. Beads were washed once in ChIA-PET wash buffer and eluted in elution buffer (10 mM Tris pH 8.0, 1 mM EDTA, 1% SDS) at 65 °C for 15 min. Reverse cross-linking was carried out at 65 °C in the presence of proteinase K. DNA was purified using phenol-chloroform-isoamyl alcohol (pH 7.9) and MaXtract High Density-2ml (QIAGEN, cat. no. 129056) and precipitated with isopropanol. Tagmentation of proximity-ligated DNA was carried out by Tn*5* transposome using the Nextera DNA Sample Preparation Kit (24) (Illumina FC-121-1030). Tagmentated DNA was purified using the Zymo Genomic DNA Clean & Concentrator kit (Zymo Research, cat. no. D4014) and fragments containing linker DNA were enriched with Dynabeads M-280 Streptavidin (Invitrogen, cat. no.11205D). Beads were washed with 2 × SSC/0.5% (wt/vol) SDS five times and twice in 1× B&W buffer (10 mM Tris pH 8.0, 1 mM EDTA, 1 M NaCl). The sequencing library was amplified using beads and the purified library was used for paired-end sequencing using MiSeq Reagent Kit v3 (600 cycle) on an Illumina MiSeq instrument.

Microarray and data processing

Both IP and input DNA were amplified using the GenomePlex complete whole-genome amplification kit (Sigma, cat. no. WGA1-50RXN), biotin-labelled and hybridized to Affymetrix GeneChip *S. cerevisiae* Tiling 1.0R Array (Sc03b_MR) according to the Affymetrix standard protocol. The CEL files were processed using rMAT[49] R package to identify enriched regions across the genome. At first, systematic biases such as probe effect were corrected by normalization. Then probe intensities were smoothed and a score was calculated for each probe using IP and input. To detect enriched regions based on the probe score, the following parameters were used; dMax = 300 (sliding window side), nProbesMin = 8 (minimum number of probes to average), method = Score (calling enriched regions based on sliding widow scores), $\log_2$ threshold = 1.5 (equal and greater than 1.5 are labelled as enriched regions). For bTMP experiments, bTMP binding 'in cells' (IP/input) was subtracted from the 'naked genomic DNA' score (IP/input) to correct for false positive binding of bTMP.

Meta-gene analysis

Meta-gene analysis was used to study the averaged enriched peak profile across all protein-coding genes (6,706 genes from SacCer 2011 annotation) or a specific set of genes upstream (−500 bases from TSS) and downstream (+500 bases from TTS) in the yeast genome. The peak scores were mapped using bedtools[50], for every base of the gene including upstream (−500 b) and downstream (+500b). The length of the gene was scaled to 1,000 bases. For scaling the ORF region to 1,000 bases, the following equation was iterated for every base across all the genes $(Z - x_i)/(y_i - x_i)) \times 1{,}000$ where x_i is the start position of the ith gene,

i = (1, 2, 3 … total genes), y_i is the end position of the ith gene, i = (1, 2, 3 … total genes), Z is base position (1, 2, 3…, $y_i - x_i$). For the average intensity plot, the IP/input values of the normalized position (1,000 ORF, −500 upstream and +500 downstream) of each gene were aggregated using median. For average gene density, the IP/input score was converted into a categorical value of either 1 or 0 based on the threshold of 1.5 (≥1.5 is 1 and <1.5 is 0) of all the normalized positions (1,000 ORF, −500 upstream and +500 downstream) of each gene and aggregated using the sum function. For visualization, average intensity and average gene density were plotted with respect to normalized ORF position. The points were smoothed using the generalized additive model (GAM) to obtain a curve using ggplot2 R package.

RNA-seq data processing

The RNA-seq data from the IonTorrent proton instrument contains approximately 25 million reads for each sample. The raw reads were filtered on the basis of quality value (−q 20 and −p 30) using the FASTX Toolkit. The filtered reads were aligned to the reference genome (SacCer 2011) using STAR aligner[51]. Aligned BAM files were used for transcript quantification (FPKM) using RSEM[52]. The gene sets were divided into three equal categories (low, medium and high expression) according to FPKM values and used to plot the supercoiling, protein and RNA–DNA hybrid profile using the meta-gene calculation mentioned above.

Histone H3 ChIP-seq data processing

The ChIP-seq data from the IonTorrent proton instrument contain approximately 15 million reads for each sample. The raw reads were filtered on the basis of quality value (−q 20 and −p 30) using the FASTX Toolkit. The filtered reads were aligned to the reference genome (SacCer 2011) using TMAP aligner. The PCR duplicates were removed from the aligned BAM files using PICARD tools. The BAM files were sorted and indexed for the peak calling using SAMtools. The bedgraph files were generated by comparing bam files of IP and input (IP read coverage/input read coverage) resulting in a ratio for every base across the whole genome using deepTools (bamCompare)[53]. Finally, peak calling was performed using the DANPOS (dpos) toolkit[54] with the IP/input threshold 1.4 (−q 1.4) where the output peaks corresponds to the individual nucleosome. The DANPOS was preferred over the MACS toolkit for the dynamic nucleosome analysis at single-nucleotide resolution.

ChIA-PET data processing

ChIA-PET data contain approximately 10 million reads with a median length of approximately 105 nucleotides. Raw reads were filtered on the basis of quality value (−q 20 and −p 30) using the FASTX Toolkit. The filtered reads were scanned for bridge linker (ACGCGATATCTTATCTGACT, AGTCAGATAAGATATCGCGT) with a maximum of two mismatches using cutadapt. The reads containing the bridge linker were aligned to the reference genome (SacCer 2011) using the bwa mem module. PCR duplicates were removed using Picard Markduplicates module. The aligned bam file was converted to a bed pair end interaction file (bedpe) for cluster generation using bedtools (bamtobed) module. PETs with less than 1 kb distance (self-ligation loops) were not considered for the PET clustering. Individual PET interactions were clustered by extending each PET by 500 bp and PETs that overlapped at both ends were clustered together as a single PET cluster[27]. PET clusters with more than or equal to 2 were considered for meta-analysis. WashU Epigenome Browser was used to visualize chromatin–chromatin interactions[55].

Tool kits

FASTX Toolkit: http://hannonlab.cshl.edu/fastx_toolkit/
TMAP Toolkit: https://github.com/iontorrent/TMAP
PICARD Toolkit: https://broadinstitute.github.io/picard/
BWA Toolkit: http://bio-bwa.sourceforge.net/

Statistics and reproducibility

All experiments were carried out with two biological replicates. To test the significance of the overlap between two replicates (supercoiling, protein and hybrid peak calls), intersect and Fisher's exact test from bedtools were used. For bedtools intersect, a minimum of 80% overlap was expected for further downstream analysis such as meta-gene plotting. The number of overlap peaks and sum of overlap bases between two sets of intervals from bedtools were visualized using VennDiagram library from R. Protein-coding genes (n = 6,706) from SacCer 2011 were used for meta-gene plotting.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All raw and processed data are available at the Gene Expression Omnibus (GEO) under the following accession numbers: GSE114410 (bTMP, RNA–DNA hybrids, Top1 protein ChIP-on-chip and RPB3 protein ChIP-on-chip); GSE114444 (RNA-seq, H3 ChIP-seq and ChIA-PET); GSE16258[47] (Top2 protein ChIP-chip, Hmo1 protein ChIP-chip and RPB3 protein ChIP-chip).

Code availability

All the custom-made scripts used for this study are available in the GitHub repository at https://github.com/adhilmd/TopologyCustomAnalysis.

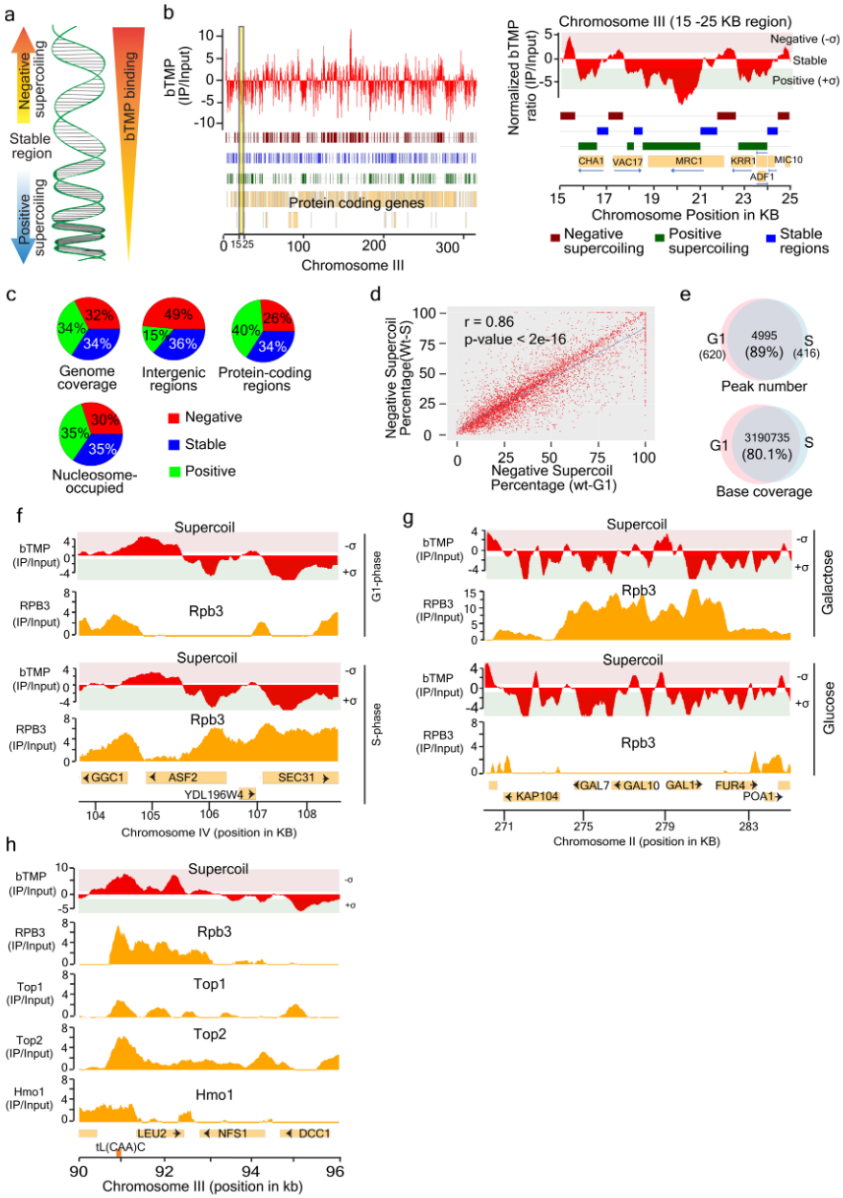
46. Thomas, B. J. & Rothstein, R. Elevated recombination rates in transcriptionally active DNA. *Cell* **56**, 619–630 (1989).
47. Bermejo, R., Katou, Y. M., Shirahige, K. & Foiani, M. ChIP-on-chip analysis of DNA topoisomerases. *Methods Mol. Biol.* **582**, 103–118 (2009).
48. Rodriguez, J., McKnight, J. N. & Tsukiyama, T. Genome-wide analysis of nucleosome positions, occupancy, and accessibility in yeast: nucleosome mapping, high-resolution histone ChIP, and NCAM. *Curr. Protoc. Mol. Biol.* **108**, 21.28.1–21.28.16 (2014).
49. Droit, A., Cheung, C. & Gottardo, R. rMAT–an R/Bioconductor package for analyzing ChIP-chip experiments. *Bioinformatics* **26**, 678–679 (2010).
50. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
51. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
52. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
53. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
54. Chen, K. et al. DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.* **23**, 341–351 (2013).
55. Zhou, X. et al. The human epigenome browser at Washington University. *Nat. Methods* **8**, 989–990 (2011).

Acknowledgements We thank J. Roca for sharing TopA-expressing plasmids, and M. Bianchi, G. Liberi and all our laboratory members for discussions. We thank Cogentech and C. Valli, M. Riboni and S. Minardi for microarray and DNA sequencing. Research was supported by grants from the Associazione Italiana per la Ricerca sul Cancro (AIRC), the European Union, MIUR, Worldwide Cancer Research, and Telethon-Italy to M.F. Y.J.A. is supported by the European Community's Seventh Framework Programme under grant agreement no. 246549 – Train 2009. N.G. is funded by the UK Medical Research Council (MR/J00913X/1; MC_UU_00007/13).

Author contributions Y.J.A. and M.F. designed the experiments, interpreted results and prepared the manuscript. Y.J.A. and M.A. performed the experiments. M.A. performed statistical and computational analysis. R.C. provided technical input and N.G. provided bTMP and technical input for supercoil analysis.

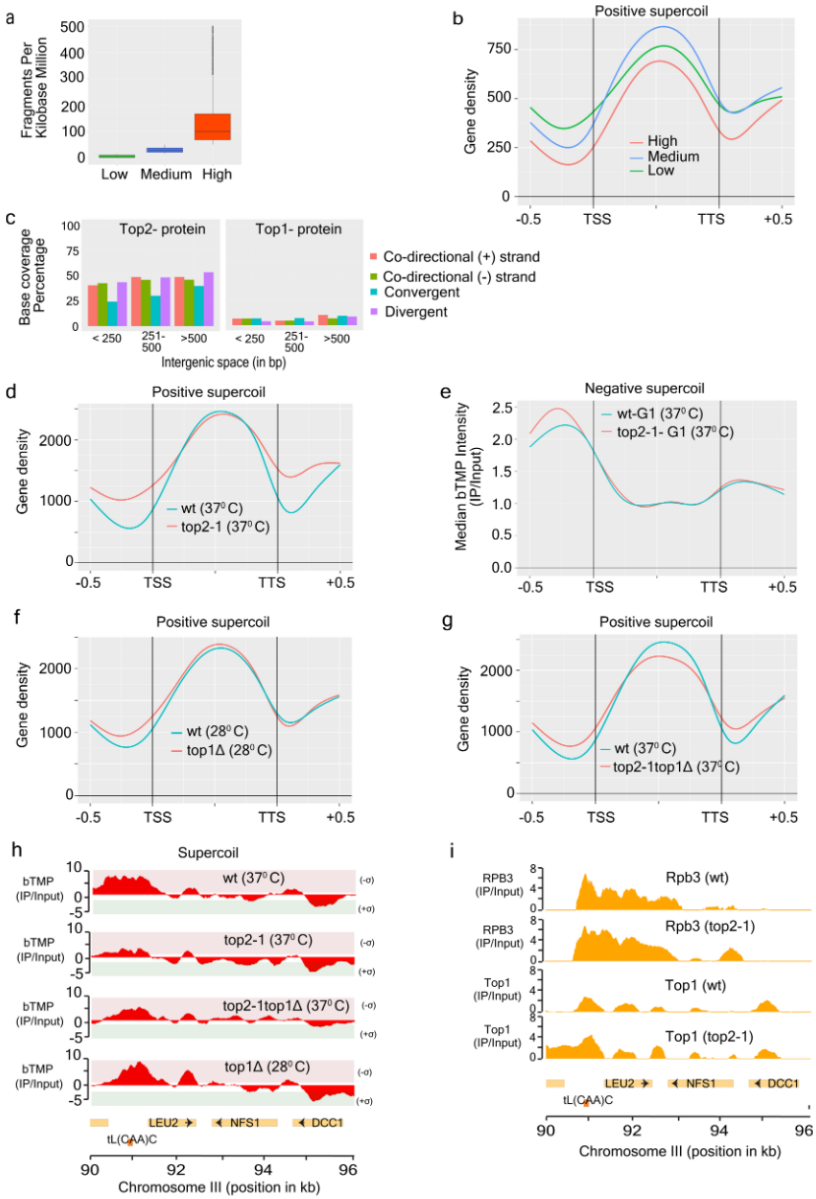
Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-1934-4.
Correspondence and requests for materials should be addressed to Y.J.A. or M.F.
Peer review information *Nature* thanks Duncan Clarke, Anne Grove and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.



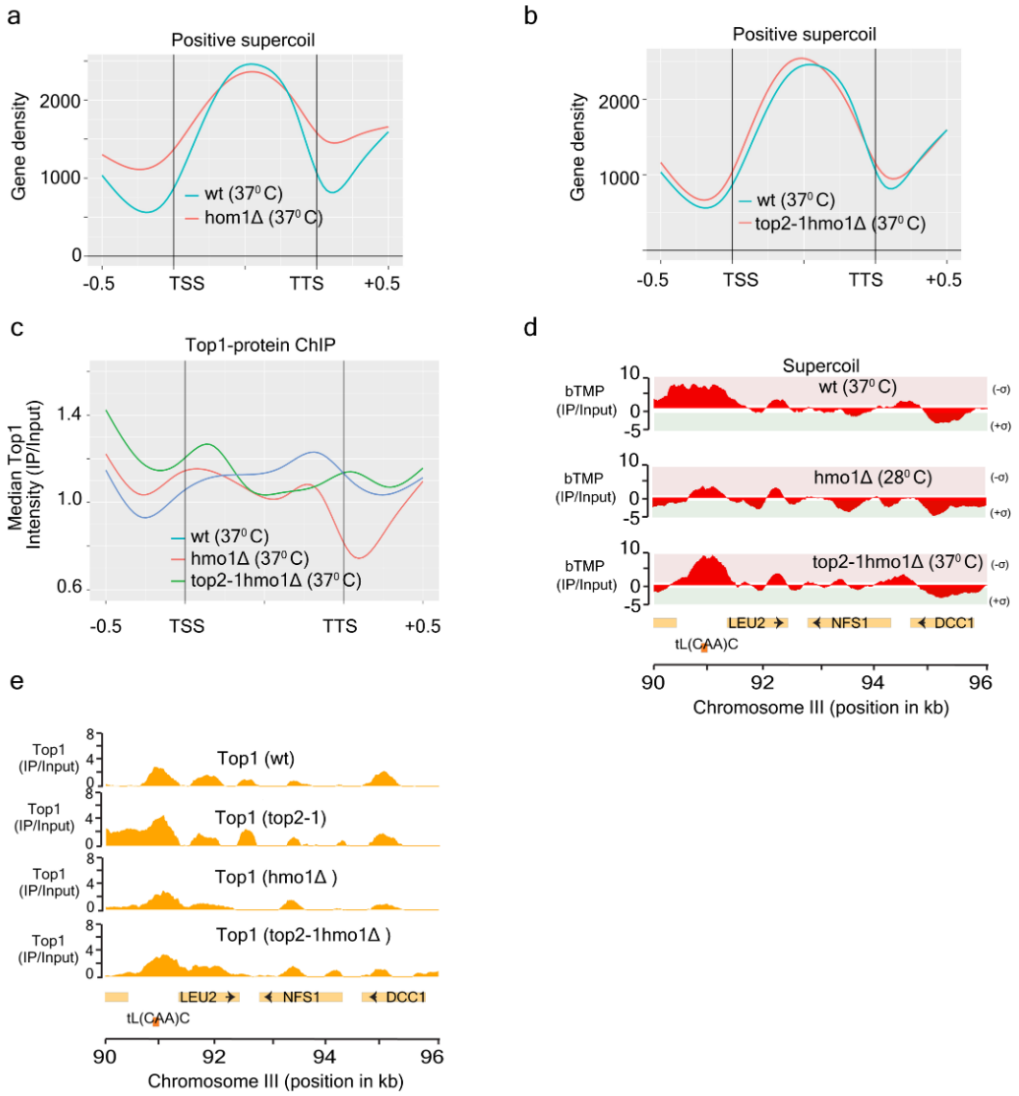
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | DNA supercoil accumulation across the genome in wild-type conditions. a, Schematic representation of bTMP binding to DNA in a supercoil-dependent manner. Based on normalized bTMP (IP/input) values, the genomic region is categorized into negative, positive and stable regions. **b**, Genome browser profile for bTMP binding on chromosome III. On the basis of peak intensities, bTMP peaks were called for negative (red), stable (blue) and positive (green) regions. Right, expanded view of chromosome III from 15 to 25 kb. For the bTMP profile, positive value peaks (normalized IP/input) that were above the threshold (+1.5) were designated as 'negative supercoil' ($-\sigma$), and negative value peaks (normalized IP/input) that were below the threshold (−1.5) were designated as 'positive supercoil' ($+\sigma$). Peaks in between the thresholds (from −1.5 to +1.5) were considered stable regions. **c**, Pie charts showing the coverage of negative, positive and stable regions based on bTMP-ChIP values plotted as percentage coverage for whole genome, intergenic regions, protein-coding regions and nucleosome-occupied regions. **d**, Gene correlation plot for negative supercoil (percentage) accumulation for wild-type cells in G1 and S phase ($n = 6{,}706$ genes; two-sided paired t-test, $P < 2 \times 10^{-16}$; Pearson correlation $r = 0.86$). **e**, Venn diagram comparison of G1 and S phase for bTMP binding with respect to peak number and base coverage. **f**, Genome browser view of the *ASF2* locus on chromosome IV, showing bTMP peaks and accumulation of Pol II (Rpb3-ChIP) in G1 and S phase. **g**, Gal genes in chromosome II (from 270 to 285 kb), depicting bTMP and Pol II (Rpb3-ChIP) binding profiles under glucose and galactose conditions in S phase. **h**, Expanded view of chromosome III from 90 to 96 kb, containing a highly active gene (*LEU2*) close to tRNA (tL(CAA)C) and two moderately expressed genes (*NFS1* and *DCC1*). Positive and negative supercoil and accumulation of Pol II (Rpb3-ChIP), Top1, Top2 and Hmo1 are shown.

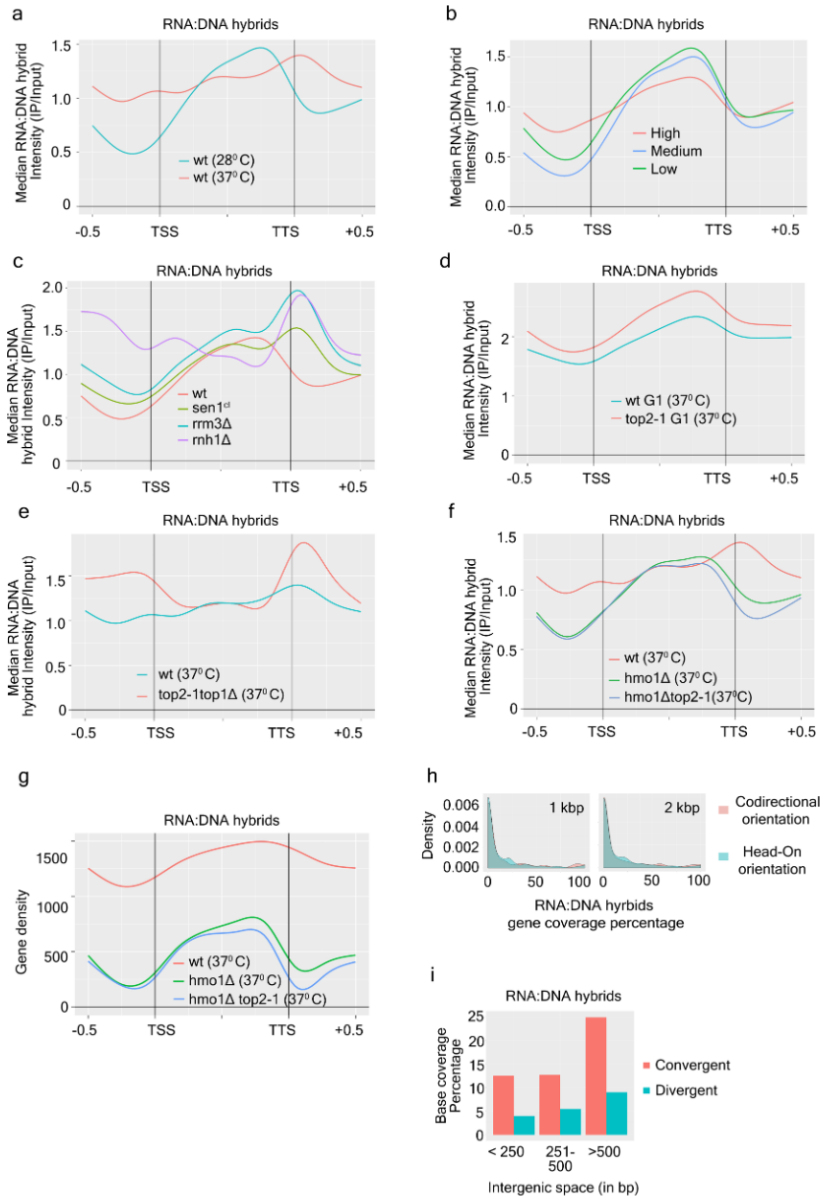


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Negative and positive supercoil distribution in DNA topoisomerase mutants. a, Pol II-coding genes were grouped into three categories; high, medium and low expression based on FPKM values from RNA-seq data (n = 6,706 genes; low = medium = high = 2,235 genes). **b**, Positive supercoil distribution in high-, medium- and low-expression genes. **c**, Base coverage percentage of accumulation of Top2 and Top1 accumulation at different intergenic spaces (<250 bp = 1,729 gene pairs, 251–500 bp = 2,224 gene pairs and >500 bp = 2,010 gene pairs) with respect to gene pairs grouped according to orientation. **d**, Accumulation of positive supercoil in wild-type and *top2-1* cells in S phase. Pol II genes are plotted against average gene density on the *y*-axis. **e**, Meta-gene plot for negative supercoil in G1 synchronized wild-type and *top2-1* cells. **f**, Accumulation of positive supercoil in wild-type and *top1Δ* cells in S phase. **g**, Accumulation of positive supercoil in wild-type and *top1Δtop2-1* cells in S phase. **h**, Genome browser profile of chromosome III from 90 to 96 kb, showing comparative bTMP binding in wild-type cells and topoisomerase mutants. **i**, Genome browser profile of chromosome III from 90 to 96 kb, showing Pol II (Rpb3-ChIP) and Top1 protein accumulation in wild-type cells and *top2-1* mutants.

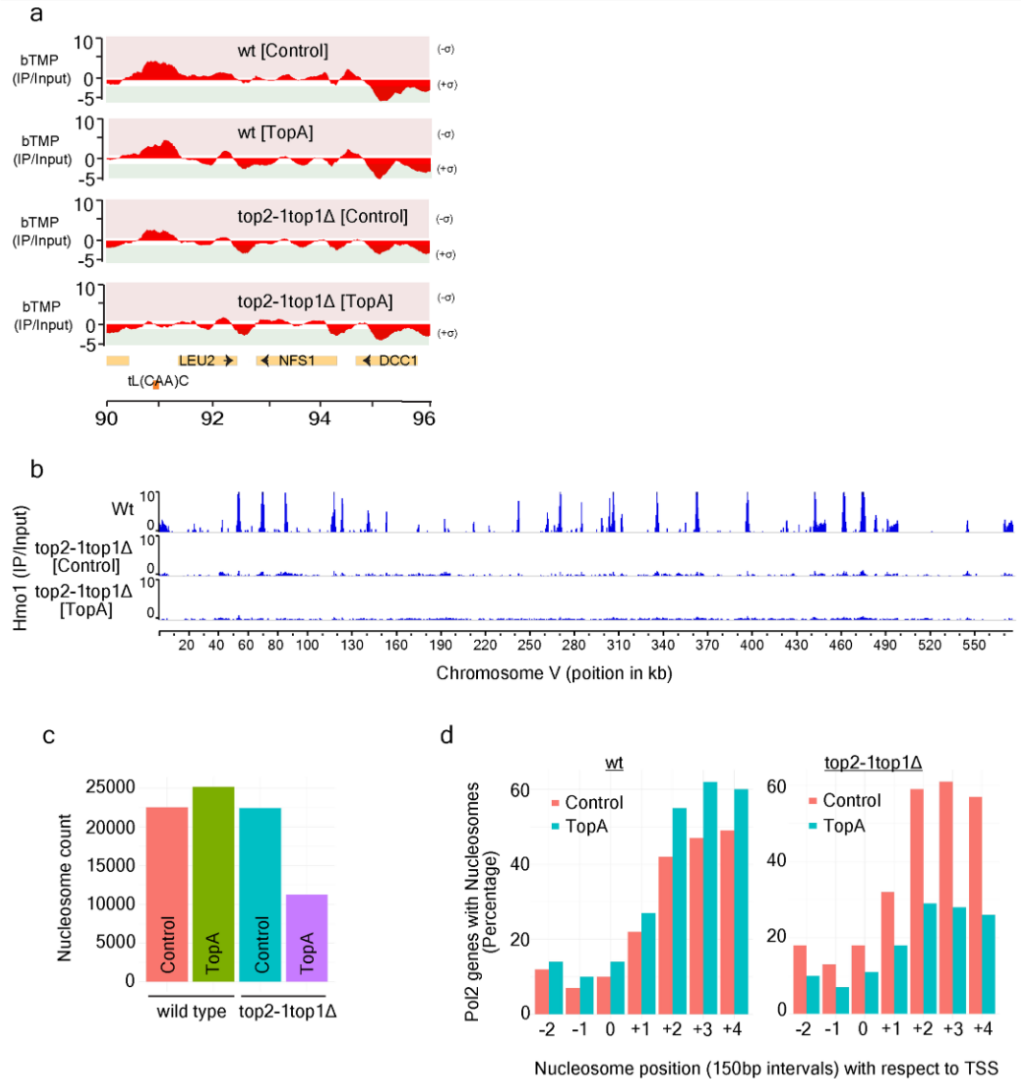


Extended Data Fig. 3 | Negative and positive supercoil accumulation in *hmo1* mutants. a, Positive supercoil accumulation in wild-type and *hmo1Δ* cells in S phase. Pol II genes are plotted against average gene density on the *y*-axis. **b**, Positive supercoil accumulation in wild-type and *hmo1Δtop2-1* cells in S phase. **c**, Comparison of the accumulation of Top1 protein (Top1–10× Flag) in wild-type, *hmo1Δ* and *hmo1Δtop2-1* cells. **d**, Genome browser profile of chromosome III from 90 to 96 kb, showing comparative bTMP binding in wild-type, *hmo1Δ* and *hmo1Δtop2-1* cells in S phase. **e**, Genome browser profile of chromosome III from 90 to 96 kb, showing accumulation of Top1 in wild-type, *hmo1Δ*, *top2-1* and *hmo1Δtop2-1* mutants.

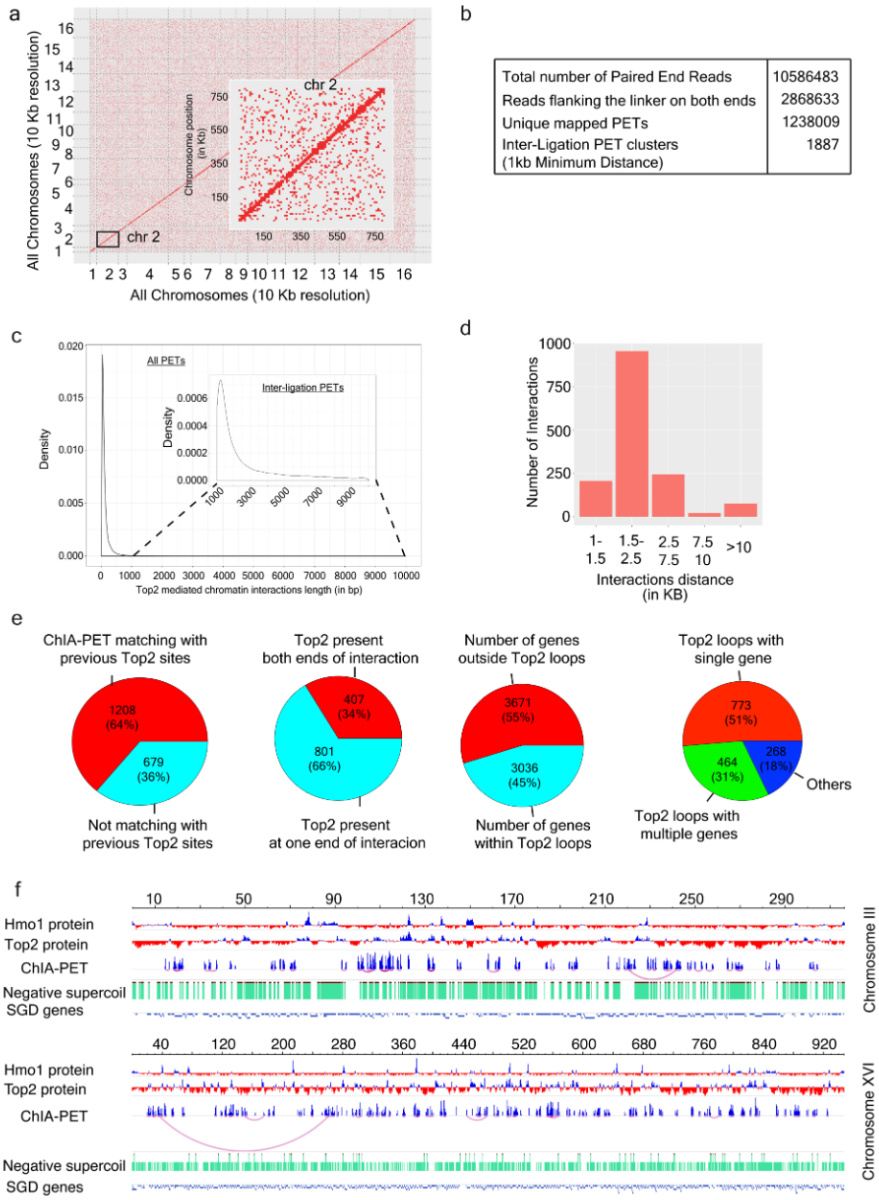


Extended Data Fig. 4 | See next page for caption.

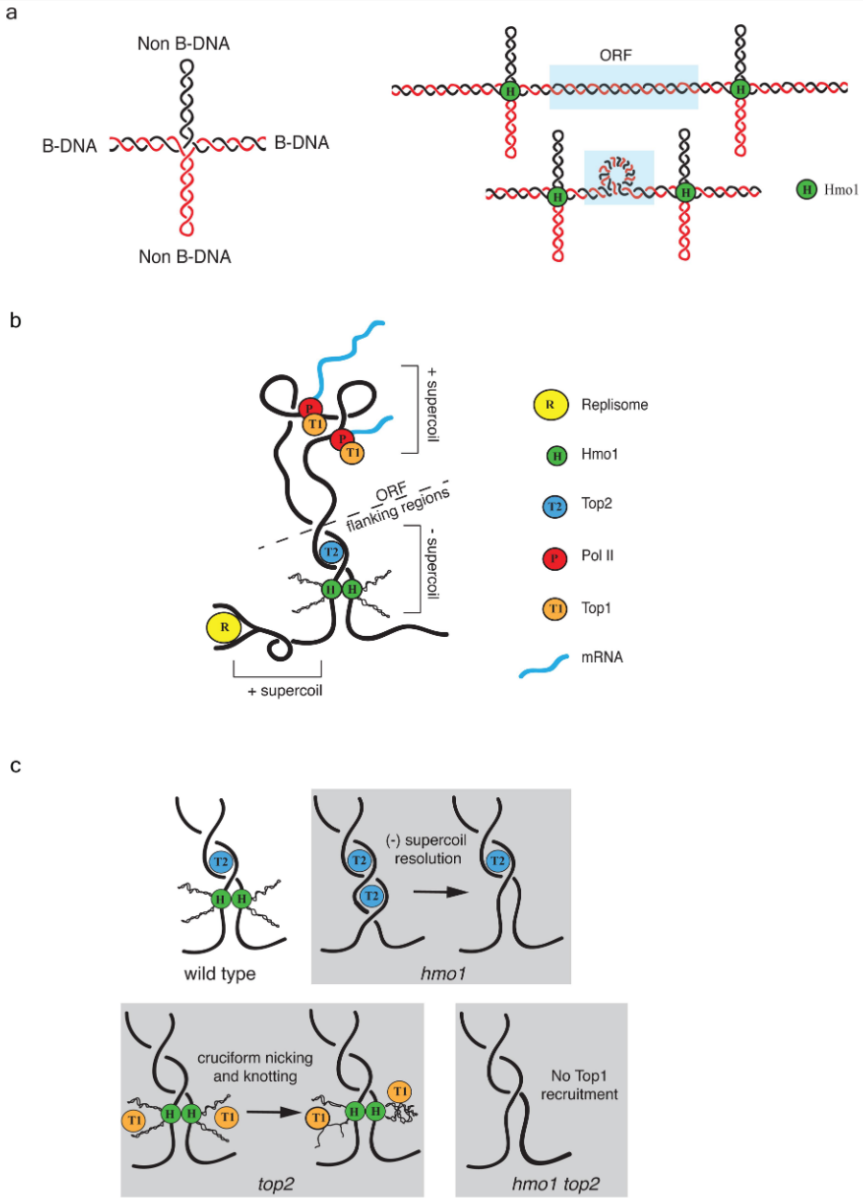
Extended Data Fig. 4 | RNA–DNA hybrid accumulation in wild-type and mutant cells. a, Meta-gene profiles for RNA–DNA hybrid comparison in wild-type cells (at 28 °C and 37 °C). **b**, Accumulation of RNA–DNA hybrids in different expression classes. Pol II genes were grouped into three categories: high-, medium- and low-expression genes based on FPKM values from RNA-seq (Extended Data Fig. 2a). **c**, Meta-gene profiles for RNA–DNA hybrids in S phase in wild-type, *rnh1Δ*, *rrm3Δ* and *sen1*^{cl} cells (a conditional lethal strain GAL::URL-HA-Sen1, which shows lethality in glucose). **d**, Meta-gene profiles for RNA–DNA hybrids in G1 synchronized wild-type and *top2-1* cells. **e**, Meta-gene profile for RNA–DNA hybrid comparison in wild-type cells and *top1Δtop2-1* double-mutant cells. **f**, Meta-gene profiles for RNA–DNA hybrid comparison in wild-type, *hmo1Δ*, and *hmo1Δtop2-1* cells. **g**, Gene density plot comparison of RNA–DNA hybrids in wild-type, *hmo1Δ* and *hmo1Δtop2-1* cells. **h**, Density plot showing the base coverage of RNA–DNA hybrids in genes with head-on or co-directional orientation with respect to replication fork. Genes within 1 kb (top, $n = 347$ genes) or 2 kb (bottom, $n = 539$ genes) of the replication origins were considered. **i**, Base coverage percentage of accumulation of RNA–DNA hybrids at different intergenic spaces (<250 bp = 1,729 gene pairs, 251–500 bp = 2,224 gene pairs and >500 bp = 2,010 gene pairs) with respect to gene pairs grouped according to orientation.



Extended Data Fig. 5 | Ablation of negative supercoil at boundaries causes topological alterations and nucleosome repositioning. a, Genome browser profile of chromosome III from 90 to 96 kb, showing comparative bTMP binding in wild-type[Control], wild type[TopA], *top1Δtop2-1* [Control] and *top1Δtop2-1* [TopA] cells after 120 min at restrictive temperature. **b**, Genome browser profile of chromosome V showing Hmo1 accumulation in wild-type, *top1Δtop2-1* [Control], and *top1Δtop2-1* [TopA] cells after 120 min at restrictive temperature. **c**, Absolute nucleosome count was derived from histone H3 ChIP-seq analysis of wild-type [Control], wild type [TopA], *top1Δtop2-1* [Control] and *top1Δtop2-1* [TopA] cells after 120 min at restrictive temperature. **d**, Nucleosome positions (150-bp intervals) in wild-type cells and *top1Δtop2-1* double mutants harbouring either control or TopA-expressing plasmids mapped with respect to TSS against Pol II gene percentage on the *y*-axis.

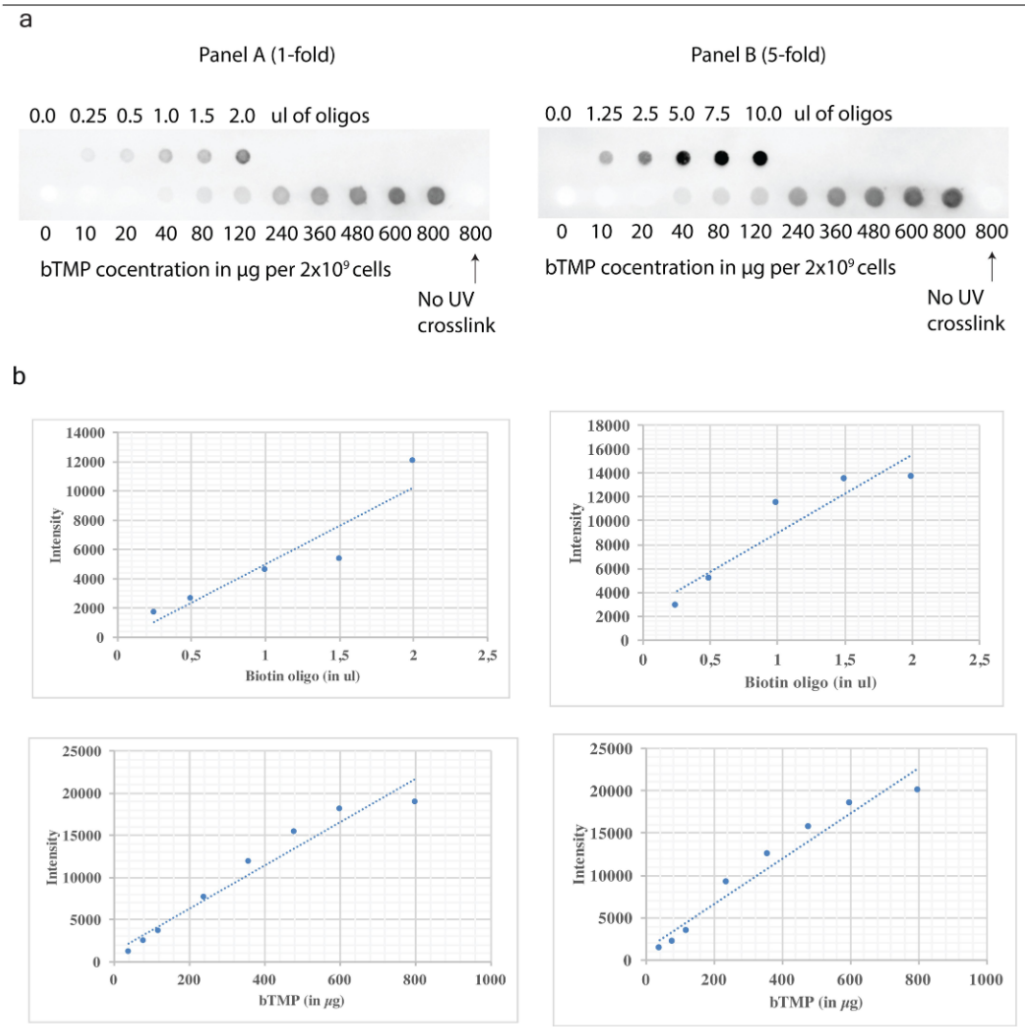


Extended Data Fig. 6 | Top2 mediates chromatin–chromatin interactions. **a**, Heat map showing PET clusters in all of the chromosomes at 10-kb resolution. Inset, enlarged view of chromosome II. **b**, Table showing the number of paired-end reads, reads with linker on both ends, unique mapped PETs and inter-ligation clusters obtained after Top2 ChIA-PET analysis. **c**, Density map showing PET density at different interaction lengths. Inlet shows enlarged view of the region from 1,000 to 10,000 bp that was considered for analysis. **d**, Bar plot showing sizewise distribution of chromatin loops. **e**, Pie chart showing number of interactions matching with previous Top2-binding sites, number of interactions containing Top2 sites on either single or both sides of the interaction, number of genes within and outside the Top2-mediated chromatin loops and number of loops containing a single gene or multiple genes. Other categories include genes partially overlapping with the loops and all other elements. **f**, Genome browser view of ChIA-PET interactions along with Top2- and Hmo1-binding sites in chromosomes III and XVI.

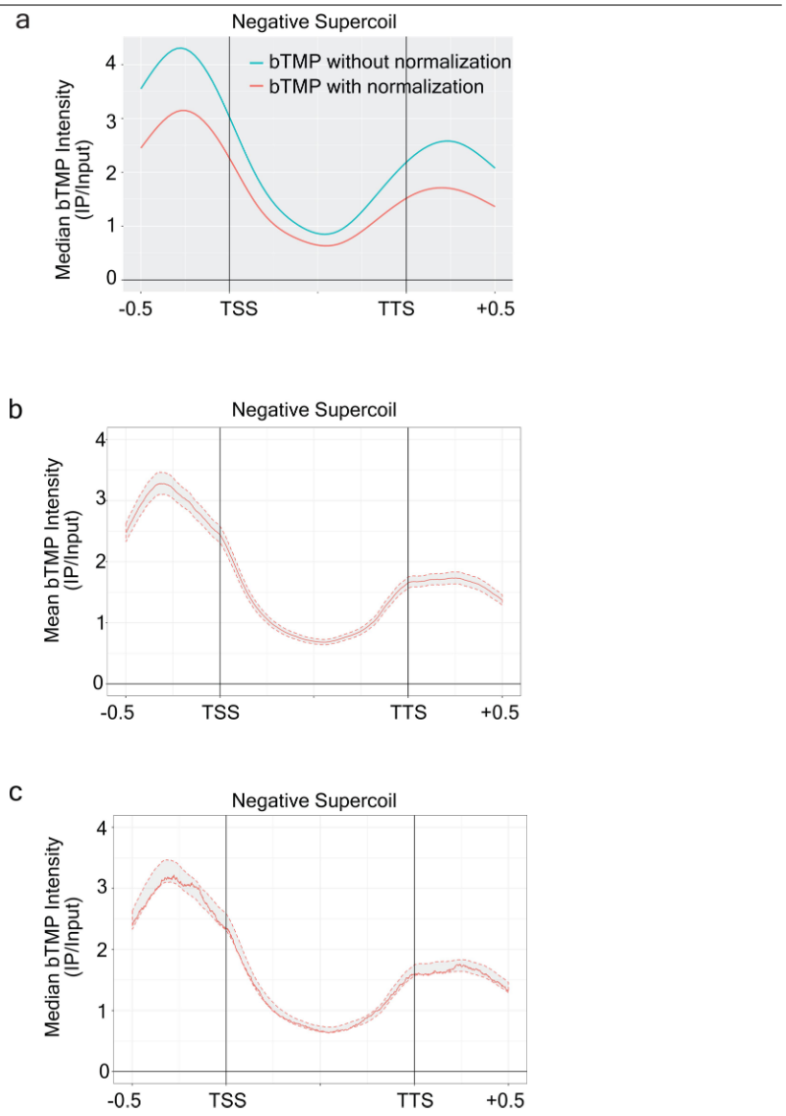


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | A model for Hmo1, Top2 and Top1-mediated topological dynamics at Pol II genes. a, Schematic representation of cruciform DNA structures arising at negatively supercoiled DNA formed by two branches in a B-DNA duplex structure (red and black) and two branches in a non-B-DNA duplex conformation (red or black). Cruciform DNA structures could form at gene boundaries and be stabilized by Hmo1. Gene looping is described in the blue area. **b**, Schematic representation of gene loop structures in S phase. Top2 associates with gene boundaries to harmonize topological transactions when transcribed genes are approached by incoming replication forks. We note that the topological dynamics described in the twin topological domain model[7] are not represented in our scheme. **c**, In the absence of Hmo1, negative supercoil would lose the cruciform conformation and become a substrate for Top2. Top2 defects would delocalize Top1 at gene boundaries. The cruciform structures stabilized by Hmo1 would then become substrates for unscheduled Top1 activity that would convert them into aberrant intermediates such as single-stranded DNA, nicks and knots. In *hmo1top2* double mutants, Top1 is not recruited at the gene boundaries, which remain in a negative supercoil context.



Extended Data Fig. 8 | bTMP titration. a, **b**, By using a biotin-labelled oligo as a reference point, we titrated in vivo binding of bTMP in yeast at different concentrations (from 0 to 800 ng). We diluted 4.16 pmol of biotin-labelled oligos into 200 µl of TE. The shown volumes of oligos were spotted onto a hybond membrane after equilibration with 1× maleic acid. Genomic DNA was isolated after UV (365 nM) cross-linking with the respective amount of bTMP and about 300 ng (**a**) or 1,500 ng (**b**) was spotted on the blot. The last sample was kept as a negative control for UV cross-linking. In Panel B, fivefold more oligo and genomic DNA was spotted compared to Panel A. Dot blot was developed with ExtrAvidin-Peroxidase antibody (Sigma-Aldrich, cat no: E2886). **b**, By measuring dot blot intensities, we estimated that 400 µg of bTMP was needed for 2×10^9 yeast cells.



Extended Data Fig. 9 | bTMP binding normalization and dispersion profile for bTMP. a, To avoid bias, normalization was performed to filter potential sequence-specific psoralen DNA binding[23]. To normalize the data, we first purified and sonicated genomic DNA and then performed the bTMP-ChIP procedure on purified DNA fragments. The correction for microarray readings was done by subtracting bTMP binding in vivo from bTMP binding on purified/sonicated genomic DNA as follows: (bTMP cells – IP/input) – (bTMP purified DNA – IP/input), which gives the normalized ratio of bTMP (bTMP – IP/input) binding. The meta ORF plot shows bTMP profiles with or without control DNA normalization of bTMP in wild-type S phase cells. **b**, **c**, Meta-gene plot showing the normalized mean bTMP ratio (**b**) and median bTMP (**c**) ratio in wild-type S phase cells. For both the plots, bTMP binding ratios for all protein-coding genes were plotted without smoothing, along with upper and lower confidence intervals ($\alpha = 0.05$ or 95% limit). Dotted lines represent upper and lower confidence intervals ($\alpha = 0.05$ or 95% limit). The confidence interval does not deviate significantly from the mean and median values.

Extended Data Table 1 | List of *S. cerevisiae* strains

	Strain	Stock Number	Genotype	Reference
1	Wt	SY2080	Mata, ade2-1, ura3-1, trp1-1, leu2-3, leu2-112, his3-11, his3-15, can1-100, GAL, PSI+, RAD5+	Lab collection
2	*top2-1*	CY8423	MATa ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3, 12 trp1-1, RAD5+, top2-1	Lab collection
3	*hmo1*Δ	CY8476	MATa, ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3, 112 trp1-1, RAD5+ ,hmo1::HIS	Lab collection
4	*top2-1hmo1*Δ	CY8475	MATa, ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3, 112 trp1-1, RAD5+, hmo1::HIS, top2-1	Lab collection
5	*top1*Δ	CY9950	MATa ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3, 12 trp1-1, RAD5+, top1::HIS	Lab collection
6	Top1-6XHis-10xFlag	CY7178	Mata, ade2-1, ura3-1, trp1-1, leu2-3, leu2-112, his3-11, his3-15, can1-100, GAL, PSI+, RAD5+, ura3::URA3/GPD-TK(7X), top1-6His10Flag (KANr)	Lab collection
7	*top2-1*,Top1-6XHis-10xFlag	CY7411	Mata, ade2-1, ura3-1, trp1-1, leu2-3, leu2-112, his3-11, his3-15, can1-100, GAL, PSI+, RAD5+, ura3::URA3/GPD-TK(7X), top1-6His10Flag (KANr), top2-1	Lab collection
8	*top2-1top1*Δ	CY10344	MATa ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3, 12 trp1-1, RAD5+, top2-1, top1::HIS	Lab collection
9	*hmo1*Δ,Top1-6XHis-10xFlag	CY15215	Mata, ade2-1, ura3-1, trp1-1, leu2-3, leu2-112, his3-11, his3-15, can1-100, GAL, PSI+, RAD5+, hmo1::Hygromycin, top1-6His10Flag (KANr)	This Study
10	*hmo1*Δ*top2-1*,Top1-6XHis-10x Flag	CY15216	Mata, ade2-1, ura3-1, trp1-1, leu2-3, leu2-112, his3-11, his3-15, can1-100, GAL, PSI+, RAD5+, top1-6His10Flag (KANr), top2-1, hmo1::Hygromycin	This Study
11	Rpb3-10X-Flag	CY15214	Mata, ade2-1, ura3-1, trp1-1, leu2-3, leu2-112, his3-11, his3-15, can1-100, GAL, PSI+, RAD5+, HIS3::BrdU-Inc, rpb3::RPB3-10X Flag-KanMX6	This Study
12	Wt [control]	CY15421	Mata, ade2-1, ura3-1, trp1-1, leu2-3, leu2-112, his3-11, his3-15, can1-100, GAL, PSI+, RAD5+ [pYEp13-LEU empty]	This Study
13	Wt [TopA]	CY15422	Mata, ade2-1, ura3-1, trp1-1, leu2-3, leu2-112, his3-11, his3-15, can1-100, GAL, PSI+, RAD5+ [pJRW13-YEptopA-pGPD-LEU]	This Study
14	*top2-1top1*Δ [control]	CY15423	MATa ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3, 12 trp1-1, RAD5+, top2-1, top1::HIS [pYEp13-LEU empty]	This Study
15	*top2-1top1*Δ [TopA]	CY15424	MATa ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3, 12 trp1-1, RAD5+, top2-1, top1::HIS [pJRW13-YEptopA-pGPD-LEU]	This Study
16	*top2-1top1*Δ,Hmo1-10X Flag, [Control]	CY15427	MATa ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3, 12 trp1-1, RAD5+, top2-1, top1::HIS, HMO1::HMO1-10X Flag [pYEp13-LEU empty]	This Study
17	*top2-1top1*Δ,Hmo1-10X Flag, [TopA]	CY15428	MATa ADE2+ CAN1+, ura3-1, his3-11,15 leu2-3, 12 trp1-1, RAD5+, top2-1, top1::HIS, HMO1::HMO1-10X Flag [pJRW13-YEptopA-pGPD-LEU]	This Study

nature research

Corresponding author(s): Yathish Jagadheesh Achar
Marco Foiani

Last updated by author(s): Oct 24, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used
Data analysis	rMAT (Droit A, 2010) (v3.33.0) STAR aligner (Dobin A, 2013) RSEM (Dewey, 2011) (v1.2.31) SAMTOOLS (v 1.9) DEEPTOOLS (Ramirez et al., 2014) (v3.2.0) DANPOS toolkit (Chen et al., 2013) (v2.2.2) BEDTOOLS (v2.29.0) FASTX Toolkit: http://hannonlab.cshl.edu/fastx_toolkit/ (v0.0.14) TMAP Toolkit: https://github.com/iontorrent/TMAP (v3.4.0) PICARD Toolkit: https://broadinstitute.github.io/picard/ (v2.14.1) MACS2 Toolkit (v2.1.2) BWA Toolkit: http://bio-bwa.sourceforge.net/ (v0.7.17-r1188) ggplot2 R package (v 3.2.1) All the custom-made scripts used for this study are made available in the GitHub repository https://github.com/adhilmd/TopologyCustomAnalysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

NCBI GEO accession ID: GSE114410 (This study) Contains bTMP, RNA:DNA hybrids, Protein ChIP-chip RAW & processed data
NCBI GEO accession ID: GSE114444 (This study) Contains RNA-seq, H3 ChIP-seq RAW, ChIA-PET RAW & processed data
NCBI GEO accession ID: GSE16258 (Bermejo et al., 2009) Contains Top2, Hmo1, RPB3 protein ChIP-chip data

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	6706 RNA pol II transcribed genes 173 replication origins
Data exclusions	No data excluded
Replication	All experiments were performed with biological replicates
Randomization	Not relevant to this study, as samples were yeast strain specific
Blinding	Not relevant to this study, as samples were yeast strain specific

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	anti-Flag (M2-antiflag) Antibody (Sigma-Aldrich cat no F3165, clone: M2, Lot#: SLBQ6349V) S9.6 Monoclonal RNA:DNA Antibody (Boguslawski et al.,1986) anti-Hstone H3 antibody (Abcam cat no 1791, Lot#:GR3236305-1) ExtrAvidin-Peroxidase antibody (Sigma-Aldrich cat no: E2886, Lot#:015M4844V, For dot blot 1:4000 dilution is used)
Validation	S9.6 Monoclonal RNA:DNA Antibody (Boguslawski et al.,1986) is validated and used in several publications including: El Hage, A., et al. (2010), Chan, Y. A., et al. (2014), Wahba, L., et al. (2016), Vanoosthuyse, V. (2018), El Hage, A. and D. Tollervey (2018)

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	NCBI GEO accession ID: GSE114444 (This study) Contains H3 ChIP-seq, Hmo1tag protein ChIP-seq, ChIA-PET RAW & processed data
Files in database submission	GSM3664900 ChIP-seq: WT-Controlplasmid-nucleosome-Input GSM3664901 ChIP-seq: WT-Controlplasmid-nucleosome-IP GSM3664902 ChIP-seq: Top2-1-Top1-Controlplasmid-nucleosome-Input GSM3664903 ChIP-seq: Top2-1-Top1-Controlplasmid-nucleosome-IP GSM3664904 ChIP-seq: WT-TopAplasmid-nucleosome-Input GSM3664905 ChIP-seq: WT-TopAplasmid-nucleosome-IP GSM3664906 ChIP-seq: Top2-1-Top1-TopAplasmid-nucleosome-Input GSM3664907 ChIP-seq: Top2-1-Top1-TopAplasmid-nucleosome-IP GSM3664908 ChIP-seq: Hmo1tag-Top2-1-Top1-Controlplasmid-protein-Input GSM3664909 ChIP-seq: Hmo1tag-Top2-1-Top1-Controlplasmid-protein-IP GSM3664910 ChIP-seq: Hmo1tag-Top2-1-Top1-TopAplasmid-protein-Input GSM3664911 ChIP-seq: Hmo1tag-Top2-1-Top1-TopAplasmid-protein-IP GSM4094803 ChIA-PET: Top2-ChIA-PET
Genome browser session (e.g. UCSC)	(i) H3 ChIP-seq (Nucleosome): http://epigenomegateway.wustl.edu/legacy/?genome=sacCer3&session=2IU7XHKkwi&statusId=1279418941 (Note: Use the Nucleosome_TopA Session: First and second track is WT-control plasmid (bed density and bedgraph), third and fourth track is WT-TopA plasmid (bed density and bedgraph), fifth and sixth track is Top2-1-Top1-Control plasmid (bed density and bedgraph), seventh and eighth track is Top2-1-Top1-TopA plasmid (bed density and bedgraph) (ii) Protein ChIP-seq: http://epigenomegateway.wustl.edu/legacy/?genome=sacCer3&session=4FQoSLC1lk&statusId=1043438118 (Note: Use the TopA_Hmo1 Session: First track is Hmo1tag-Top2-1-Top1-Controlplasmid and Second track is Hmo1tag-Top2-1-Top1-TopAplasmid) (iii) ChIA-PET: http://epigenomegateway.wustl.edu/legacy/?genome=sacCer3&session=arHkgZiCZP&statusId=302725081 (Note: Use the ChIA-PET-Data Session: First track is Hmo1 protein binding regions, Second track is Top2 protein binding regions, Third track is ChIA-PET data for Top2-10X Flag and Fourth track is the interaction points for ChIA-PET Top2-10X Flag) No longer applicable for final submission.

Methodology

Replicates	One biological replicate for each sample is generated and analyzed for consistency. (Replicates were not yet deposited in the GEO)
Sequencing depth	(i) H3 ChIP-seq (Nucleosome): Average Reads: ~15 Million Reads, Average Read length: ~180bp, Average Mapped Reads: ~8 Million Reads and single end reads (Ion Torrent Platform) (ii) Protein ChIP-seq: Average Reads: ~15 Million Reads, Average Read length: ~180bp, Average Mapped Reads: ~8 Million Reads and single end reads (Ion Torrent Platform)
Antibodies	(i) H3 ChIP-seq (Nucleosome): anti-Hstone H3 antibody (Abcam cat no 1791) (ii) Protein ChIP-seq: anti-Flag (M2-antiflag) Antibody (Sigma-Aldrich cat no F3165) (iii) ChIA-PET: anti-Flag (M2-antiflag) Antibody (Sigma-Aldrich cat no F3165)
Peak calling parameters	(i) H3 ChIP-seq: The raw reads were filtered based on the quality value (-q 20 and –p 30) using FASTX Toolkit. The filtered reads were aligned to the reference genome (SacCer 2011) using TMAP aligner. The PCR duplicates were removed from the aligned BAM files using PICARD tools. The BAM files were sorted and indexed for the peak calling using SAMTOOLS. The bedgraph files were generated by comparing bam files of IP and Input (IP read coverage/Input read coverage) result in the ratio for every base across the whole genome using DEEPTOOLS (bamCompare) (Ramirez et al., 2014). Finally, peak calling was performed using DANPOS (dpos) toolkit (Chen et al., 2013) with the IP/Input threshold 1.4 (-q 1.4) where the output peaks corresponds to the individual nucleosome. The DANPOS was preferred over MACS toolkit for the dynamic nucleosome analysis at single nucleotide resolution. (ii) Protein ChIP-seq: The raw reads were filtered based on the quality value (-q 20 and –p 30) using FASTX Toolkit. The filtered reads were aligned to the reference genome (SacCer 2011) using TMAP aligner. The PCR duplicates were removed from the aligned BAM files using PICARD tools. Finally, MACS2 tool is used for peak calling with the following parameters (-f BAM --gsize=1.21e+7 -n ctrl-A -B -p 0.01 --nomodel --extsize 200 --broad).
Data quality	The peak fold enrichment were in the range of 1.5 to 3.0

Software	SAMTOOLS (v 1.9) DEEPTOOLS (Ramirez et al., 2014) (v3.2.0) DANPOS toolkit (Chen et al., 2013) (v2.2.2) BEDTOOLS (v2.29.0) FASTX Toolkit: http://hannonlab.cshl.edu/fastx_toolkit/ (v0.0.14) TMAP Toolkit: https://github.com/iontorrent/TMAP (v3.4.0) PICARD Toolkit: https://broadinstitute.github.io/picard/ (v2.14.1) MACS2 Toolkit (v2.1.2) BWA Toolkit: http://bio-bwa.sourceforge.net/ (v0.7.17-r1188)