Elucidating the role of the Smc5/6 complex in genome organization and repair

Shamayita Roy

A thesis submitted in partial fulfillment of the requirements for the

Doctorate in Philosophy degree

in Cellular and Molecular Medicine

Department of Cellular and Molecular Medicine
Faculty of Medicine
University of Ottawa

List of Publications and contributions

Manuscript #1:

Shamayita Roy**,** Arvin Zaker, Arvind Mer, Damien D'Amours, Large-scale phenogenomic analysis of human cancers uncovers frequent alterations affecting SMC5/6 complex components in breast cancer, NAR Cancer, Volume 5, Issue 3, September 2023, zcad047, https://doi.org/10.1093/narcan/zcad047

[SR and DD conceived and designed yeast experiments; AZ and AM performed all bioinformatics analyses; SR created yeast strains, assessed their viability and DNA damage sensitivity; SR and AZ, prepared figures; AZ and AM wrote all the bioinformatics sections of the paper; SR and DD wrote the sections of the paper pertaining to yeast experiments.]

License Number: 5831060990014

License date: Jul 16, 2024

Licensed Content Publisher: Oxford University Press

Licensed Content Publication: NAR Cancer

Manuscript #2:

Shamayita Roy, Hemanta Adhikary, Sarah Isler & Damien D'Amours, The Smc5/6 complex counteracts R-loop formation at highly transcribed genes in cooperation with RNase H2.

Manuscript submitted and in revision at ***eLife***.

[SR and DD conceived and designed the experiments; SR created yeast strains, assessed their viability and DNA damage sensitivity; SR performed all the microscopy and DRIP experiments. HA performed all the biochemical experiments; SI performed the pull-down experiments; SR, HA and DD analyzed the data and prepared figures; SR, HA and DD wrote the paper.]

Manuscript #3 (review article):

Shamayita Roy**,** Hemanta Adhikary, Damien D'Amours, The SMC5/6 complex: folding chromosomes back into shape when genomes take a break, *Nucleic Acids Research*, Volume 52, Issue 5, 21 March 2024, Pages 2112–2129, https://doi.org/10.1093/nar/gkae103

Included in **Chapter I Introduction** of this thesis: From section "CHROMATIN ORGANIZATION AND SMC-FAMILY ENZYMES" to "CONCLUSION".

License Number: 5831070360604

License date: Jul 16, 2024

Licensed content publisher: Oxford University Press

Licensed content publication: Nucleic Acids Research

Manuscript #4 (Co-Author Appendix):

Serrano, D., Cordero, G., Kawamura, R., Sverzhinsky, A., Sarker, M., Roy, S., Malo, C., Pascal, J.M., Marko, J.F., and D'Amours, D. (2020). The Smc5/6 Core Complex Is a Structure-Specific DNA Binding and Compacting Machine. Mol Cell 80, 1025-1038.e5. 10.1016/j.molcel.2020.11.011.

License Number: 5831061462971

License date: Jul 16, 2024

Licensed Content Publisher: Elsevier

Licensed Content Publication: Molecular Cell

Abstract

Effective repair and maintenance of genetic material during cell proliferation is vital for cell survival. In this context, structural maintenance of chromosome (SMC) proteins play crucial and well-defined functions in the maintenance of genomic integrity. Specifically, the Smc5/6 complex has multifarious roles in folding and repairing chromosome segments to preserve genomic stability. How the Smc5/6 complex performs such roles has been at the forefront of research efforts for decades. In this thesis, I explored the molecular basis for the extensive roles of the Smc5/6 complex on damaged chromatin and its impact on cellular health and human diseases. An in-depth bioinformatic analysis of mutations affecting the human SMC5/6 complex in cancer genomics cohorts revealed that subunits of the SMC5/6 complex are frequently altered in several types of cancers. I introduced a large number of SMC5/6 complex mutations identified in human cancers into the budding yeast *Saccharomyces cerevisiae* and demonstrated that most of these mutations are associated with a loss of Smc5/6 complex activity. My results suggest a direct link between loss of SMC5/6 complex activity and cancer development. Next, I investigated the nature of endogenous substrates bound by the SMC5/6 complex on chromosomes. I found that R-loop structures, a common bi-product of chromosomal transactions, are recognized by the SMC5/6 complex *in vivo*. To identify whether Smc5/6 complex promotes genomic integrity by controlling R-loop formation, I explored the possibility that the Smc5/6 complex participates in the degradation of R-loop structures. Importantly, I demonstrated that the Smc5/6 complex promotes the degradation of R-loops by RNase H2 enzyme at highly transcribed genes and telomeres. Finally, I identified important interfaces within the subunits of the Smc5/6 complex that are responsible for its repair activity and showed that mutations in these regions may lead to chromosome breakage syndromes in humans. Taken together, this thesis reveals novel functions of the Smc5/6 complex and how misregulation of this enzyme impacts human health and disease states.

List of Abbreviations

3D = three dimensional

4NQO = 4-nitroquinoline 1-oxide

AESBF = 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride

AID = activation-induced cytosine deaminase

ALT = alternative lengthening of telomeres

ARD = ankyrin repeat domain

ARM = armadillo

ATP = Adenosine triphosphate

BARD1 = BRCA1-associated RING domain protein 1

B cells = Bursa-derived cells

BER = base excision repair

BLM = Bloom syndrome protein

BSA = bovine serum albumin

° C = degree Celsius

CNA = copy-number alteration

CRISPR-Cas9 = clustered regularly interspaced short palindromic repeats and CRISPR-associated protein

Cy3 = Sulfo-Cyanine3

DAPI = 4',6- diamidino-2-phenylindole

DDR = DNA damage response

DNA = Deoxyribonucleic acid

DRIP = DNA:RNA Immunoprecipitation

DSB = double strand break

dsDNA = double stranded DNA

DTT = Dithiothreitol

EBV = Epstein–Barr virus

EDTA = Ethylenediaminetetraacetic acid

EM = electron microscopy

EMSA = electrophoretic mobility shift assay

FDR = false discovery rate

FP = false positive

FPLC = Fast Protein Liquid Chromatography

γH2A = gamma Histone 2A

ATP, [γ-^{32}P] = Adenosine triphosphate, labeled on the gamma phosphate group with ^{32}P

GFP = green fluorescent protein

GLM = generalized linear model

GSEA = gene-set enrichment analysis

GST = Glutathione S-transferase

HBV = hepatitis B virus

HBx = hepatitis B virus regulatory protein X

HEPES = 4-(2-Hydroxyethyl)-1-piperazine ethanesulfonic acid

HIV = human immunodeficiency virus

HJ = Holliday junction

HPV = human papilloma virus

HR = homologous recombination

hrs = hours

HTH = helix-turn helix

HU = hydroxyurea

IR = ionizing radiation

KCl = Potassium chloride

KITE = kleisin-interacting tandem winged-helix element

KM = Kaplan-Meier

KPO_4 = Potassium phosphate

KSHV = Kaposi's sarcoma herpesvirus

LFC = log fold change

LICS = lung disease immunodeficiency and chromosome breakage syndrome

MES = 2-(N-morpholino) ethanesulfonic acid

METABRIC = Molecular Taxonomy of Breast Cancer International Consortium

$MgCl_2$ = Magnesium chloride

μm = micrometer

MMR = mismatch repair

MMS = methyl methanesulfonate

MOPS = morpholinepropanesulfonic acid

MSigDB = Molecular Signatures Database

MVH = mosaic variegated hyperploidy

NaCl = Sodium Chloride

NER = nucleotide excision repair

NHEJ = non-homologous end joining

Ni-NTA = Nickel Nitriloacetic acid

NP-40 = Nonidet P40

NSE = Non-SMC element

O.D. = optical density

P.C.R = Polymerase Chain Reaction

PDB = Protein Data Bank

PML-NB = promyelocytic leukemia nuclear bodies

POL2 = Polymerase II

psi = pounds per square inch

PTM = post translational modification

qPCR = quantitative Polymerase Chain Reaction

rDNA = ribosomal DNA

RF = replication fork

RNA = Ribonucleic Acid

RNase H = Ribonuclease H

ROS = reactive oxygen species

RPA = Replication protein A

SCI = sister chromatids intertwinement

SDS-PAGE = Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis

SEM = standard error of mean

SIMC1 = SUMO Interacting Motifs Containing 1

SLF1 = SMC5/6 complex localization factors 1

SLF2 = SMC5/6 complex localization factors 2

SMC = structural maintenance of chromosomes

ssDNA = single stranded DNA

STR = SGS1-TOP3-RMI1

SV = structural variation

TADS = topologically associating domains

TBS = Tris-buffered saline

TCA = Trichloro acetic acid

TCGA = The Cancer Genome ATLAS

TERRA = telomeric repeat-containing RNA

TEV = Tobacco etch virus

TLS = translesion DNA synthesis

TONSL = Tonsoku-like protein

TOP = Topoisomerase

TP = true positive

UV = ultraviolet light

VAF = variation allele frequency

VUS = variants of unknown significance

WH = Winged helix

YPD = yeast extract peptone glucose

Acknowledgements

I am extremely fortunate to have incredible support from a few people who made the journey of my PhD effortless and enjoyable. First and foremost, my supervisor, Damien D'Amours. Thank you for providing continuous inspiration and support throughout my PhD. Your guidance and encouragement have helped me to believe in myself and navigate this journey with ease. It has been a pleasure to have you as my mentor.

I also extend my gratitude to my thesis committee members Dr. Michael Downey, Dr. Laura Trinkle Mulcahy and Dr. John Copeland for their valuable feedback which helped me immensely during my PhD research. I would like to thank the University of Ottawa for financial support. I would also like to thank all my past and present collaborators whose invaluable help made this thesis a reality. Notably, Dr. Arvind Mer and Arvin Zaker for their work on the computational portion of Chapter II and Dr. Hemanta Adhikary and Sarah Isler for their help with the biochemical and other experimental help in Chapter III. Also, I would like to extend my gratitude to Dr. Diego Serrano who was immensely helpful during the conception of my study shown in the Appendix of this thesis.

I will be forever indebted to the members of the D'Amours laboratory for their words of encouragement and wisdom. You were my home away from home for the past 6 years. Thank you to Catherine Malo for helping me to get started in the lab, for helping me set up all my experiments in the first year and a strong source of moral support during the crazy "western blot days". Thank you to Annahat for being my true confidant. Thank you to Laurence and Alyssa for being the best co-passengers on this journey. Special thanks to the 'big brother' Hemanta for being the calming presence in the middle of all the chaos.

This work would have been impossible without the constant support from my family members. Thank you to my parents, Shyamal Roy and Sharmila Roy without whose support, I would not have taken the first flight to Canada to begin with. Thanks to my brother for silently being my biggest cheerleader. Also, a special thank you to my mother- and father-in-law, for their non-stop motivation.

Finally, I would not have done a PhD if it was not for my dear husband Partha. I dedicate this thesis to you. Thank you for being there, come rain or shine.

Table of Contents

List of figures

List of supplementary figures

List of tables

List of supplementary tables

Chapter I: General Introduction

1.1 STRUCTURAL ORGANIZATION OF EUKARYOTIC GENOMES

The organization of chromatin in space is one of the key determinants of the functional state of genes and genomes, linking nuclear level activities to overall organismal health and well-being. Hence, a comprehensive understanding of the structural modifications that modulate the spatial organization of genomes has an immense fundamental and applied value in biology. The eukaryotic genome constantly undergoes large-scale structural modifications during its life cycles. With the progression of cell cycle events, the genome changes its structure from an open decondensed state to an intermediate structure in interphase and finally forms fully condensed chromosomes in mitosis and meiosis (Dekker & Mirny, 2016; Golloshi *et al*, 2017). The last several decades of research and the advent of cutting-edge genomics, biochemistry, and biophysical techniques has revealed that open or accessible chromatin generally promotes DNA transactions while densely folded chromatin significantly limits enzyme mobility and access to DNA substrates (Bell *et al*, 2011; Shi *et al*, 2022). An intricate network of factors help eukaryotic organisms to finely tune the organization of chromatin in space to enhance the efficiency, fidelity, and regulation of vital cellular processes such as gene transcription, DNA repair and genome segregation (Cardozo Gizzi *et al*, 2019).

1.2 GENOMIC INSTABILITY DUE TO DNA DAMAGE

One major challenge that cells have to endure in order to maintain the proper structural organization of chromosomes is DNA damage. Cells are constantly exposed to extensive exogenous and endogenous DNA lesions that compromise the integrity of their genomes (Huang

& Zhou, 2020; Carusillo & Mussolino, 2020; Norbury & Hickson, 2001). Furthermore, errors during DNA replication can jeopardize genome stability. Independent of the form of the DNA damage, the most common effect of DNA lesion is the activation of DNA damage checkpoints and cell cycle arrest, allowing time for the repair machinery to restore DNA integrity (Huang & Zhou, 2020; Pedroza-Garcia *et al*, 2022). Erroneous DNA repair processes can lead to mutations and chromosomal aberrations, two types of alterations that have been commonly linked to premature ageing and cancer (Schumacher *et al*, 2021). Furthermore, studies have shown that many of the human diseases, such as cancers, arise from germ-line mutations in DNA damage repair genes (Xu *et al*, 2021; Sharma *et al*, 2020; Sherwood *et al*, 2023).

What are the sources of DNA damage? Endogenous DNA damage can be an outcome of errors in proofreading during replication, deamination as well as depurination/depyrimidination of bases. Additionally, metabolic bi-products of normal cellular processes can result in oxidative stress due to the formation of reactive oxygen species (ROS), generating a variety of DNA lesions, such as oxidized bases and strand breaks (Hahm *et al*, 2022; Degtyareva *et al*, 2013). Moreover, replication and transcription errors can often lead to ribonucleotide incorporation or excess amount of RNA-DNA hybrid formation respectively which can interfere with normal DNA transactions (Appanah *et al*, 2020; Li *et al*, 2023; Petermann *et al*, 2022). On the other hand, exogenous DNA damage is caused by external stressors such as ionizing radiation (IR), ultraviolet light (UV), hydrolysis or thermal disruption, mutagenic chemicals and other similar DNA damaging agents that directly modify nucleic acid structure (Hakem, 2008; De Bont & van Larebeke, 2004; Lindahl & Barnes, 2000; Tubbs & Nussenzweig, 2017; Potenski & Klein, 2014).

Upon DNA damage, cells induce DNA damage response (DDR) which involves DNA damage recognition followed by signal transduction, cell cycle control and finally DNA repair to resolve the lesion. These complex DDR pathways include dedicated 'sensor' proteins to recognize DNA damage and 'transducer' proteins that recruit subsequent 'effector' proteins responsible for cell cycle arrest and DNA repair (Liu *et al*, 2016; Misteli & Soutoglou, 2009).

Under normal physiological conditions, there are six major DNA repair pathways to combat DNA lesions: base excision repair (BER) that fixes damaged DNA bases or single-strand DNA breaks resulting from deamination or hydroxylation of bases, mismatch repair (MMR) for mis-paired nucleotides, nucleotide excision repair (NER) which combats a variety of helix-distorting DNA lesions, translesion DNA synthesis (TLS) to bypass DNA adducts, and finally nonhomologous end joining (NHEJ) and homologous recombination (HR) to repair double stranded DNA breaks (DSBs) (Tian *et al*, 2015; Huang & Zhou, 2020; Norbury & Hickson, 2001).

It is becoming increasingly clear that a close collaboration between the DDR pathways and chromosome organization is required to uphold faithful maintenance of genomic integrity. Last few decades of research have identified the major molecular players that modulate both of these aspects of genomic maintenance – the roles of which are discussed below.

1.3 CHROMATIN ORGANIZATION AND SMC-FAMILY ENZYMES

Changes in chromatin state is a complex affair typically requiring multi-step processes that involve nucleosomes, post-translational modifications (PTMs) of histones, as well as chromatin

remodeling enzymes (Bell *et al*, 2011; Shi *et al*, 2022). This paradigm represents the prevalent view of how chromatin organization is regulated in interphase cells. However, recent research has revealed that large-scale changes in genome organization can be imparted through completely different processes, often independently of canonical chromatin/histone modifying enzymes. One of these new processes is based on the unique capacity of a family of enzymes to create topological links that connect distant regions of chromosomes, thus generating DNA loops of varying sizes and complexity. The key enzymes responsible for this level of chromatin organization are the structural maintenance of chromosomes (SMC) family complexes (Kim *et al*, 2023). While these complexes have well-documented roles in mitosis, they also play crucial roles in interphase chromatin folding, regulation as well as DNA damage repair. What makes these enzymes interesting is their unique ability to regulate chromatin folding by themselves, independently of nucleosomes and most other enzymes typically associated with chromatin compaction processes (Shintomi *et al*, 2017). As such, SMC complexes represent a fascinating alternative to canonical chromatin modification pathways to regulate the repair and homeostasis of genes and genomes.

SMC family proteins are large ATPases known to play paramount roles in the spatial and temporal organization of chromosomes in all domains of life. Eukaryotic genomes typically encode six evolutionarily conserved SMC proteins, while prokaryotes usually express a single SMC family member. To perform any of their designated roles, these SMC proteins need to assemble into large holoenzyme complexes in two distinct steps. First, SMC monomers associate as pairs with cognate partners (see example in Figure 1.1A) to create specific heterodimers: SMC1/SMC3,

SMC2/SMC4 and SMC5/SMC6. Second, SMC heterodimers associate with additional non-SMC elements (NSEs) to generate three different and fully functional holoenzyme complexes: cohesin, condensin and the SMC5/6 complex. The relative positioning of NSE elements within SMC complexes together with multiple bivalent interactions among complex subunits (Figure 1.1) generates the prototypical ring-like architecture of SMC complexes and a functional identity that is unique to each holoenzyme (reviewed in refs. (Bürmann & Löwe, 2023; Hoencamp & Rowland, 2023; Kim *et al*, 2023)).

The SMC1 and SMC3 family members form the heterodimer at the core of the cohesin complex, and together with additional NSEs, Scc1/Mcd1/RAD21 and Scc3/STAG/SA, holds the replicated sister chromatids together. This process ensures error free segregation of chromatids during mitosis (Guacci *et al*, 1997; Michaelis *et al*, 1997). Cohesin also participates in the establishment of complex chromatin structures, called topologically associating domains (TADs), to regulate gene expression and DNA replication during interphase (Wutz *et al*, 2017; Hansen *et al*, 2018; Emerson *et al*, 2022). Likewise, SMC2 and SMC4 family members associate with CAP-D/G/H-family subunits (Ycg4/Cnd1, Ycg1/Cnd3 and Brn1/Cnd2 in yeast, respectively) to form the condensin holoenzyme, the main effector of chromosome condensation during mitosis (Hirano *et al*, 1997).

The third and the most enigmatic SMC holoenzyme, the SMC5/6 complex, demonstrates a unique subunit composition and shape, thanks to an unusual number of NSE components and atypical positions of some of its subunits within the complex (Peng & Zhao, 2023) (Figure 1.1). The non-

SMC subunits of the SMC5/6 holoenzyme include Nse1-6 in yeasts and NSMCE1-4 in mammals (note that Nse2 is also known as Mms21 in budding yeast; (Fujioka *et al*, 2002; Hazbun *et al*, 2003; McDonald *et al*, 2003; Pebernard *et al*, 2004, 2006; Hu *et al*, 2005; Potts & Yu, 2005; Sergeant *et al*, 2005; Zhao & Blobel, 2005; Taylor *et al*, 2008)). Humans also co-express SMC5/6 complex localization factors 1 and 2 (SLF1 and SLF2), believed to be functional homologues of yeast Nse5 and Nse6 respectively, but it is unclear if these subunits are constitutive members of the human complex (Figure 1.1C) (Räschle *et al*, 2015). Therefore, when it comes to non-SMC partners, the SMC5/6 complex has a significantly greater number of non-SMC subunits than cohesin or condensin complexes. For brevity and clarity, we will use below the "NSE" nomenclature when discussing non-SMC subunits of the SMC5/6 complex, as well as the human notation convention for all proteins discussed hereafter in Chapter I.

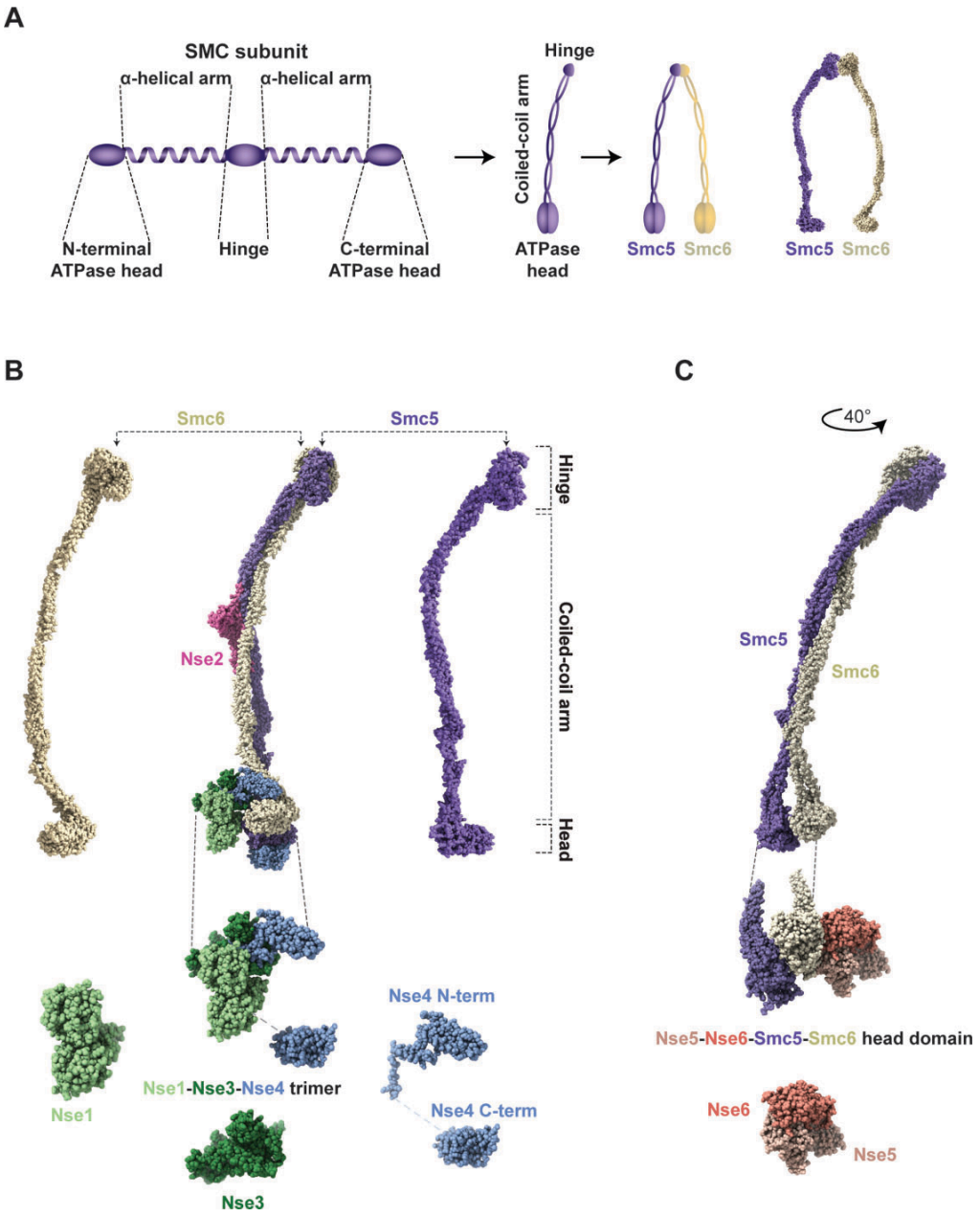


Figure 1. 1 The architecture of the SMC5/6 complex. (A) Schematic representation of the functional domains, folding, and dimerization states of SMC5 and SMC6 proteins. Shown on the left is a cartoon representation of the domain organization of a typical SMC protein in its unfolded state. The middle section of the panel depicts the same SMC monomer in its mature state. Specifically, the newly synthesized SMC protein initially fold back on itself (in a manner analogous to a conventional hairpin) to create a functional coiled-coil arm and ATPase head domain. This allows the folded/mature protein to dimerize with another SMC family member using their respective hinge domains. The right side of the panel shows the SMC5/6 dimer in cartoon and structural representations. (B) Cryo-EM composite structure and molecular graphics representation of the SMC5/6 complex and its individual subunits edited in UCFC ChimeraX (Pettersen *et al*, 2021). The hinge, coiled-coil arm and ATPase head domains are labelled on the

right-hand side of the SMC5 structure. Individual subunits/subcomplexes are shown in different color schemes to facilitate visualization; SMC5, violet; SMC6, wheat/beige; NSE1, light green; NSE2, pink; NSE3, dark green; NSE4, blue; NSE5, peach; and NSE6, maroon (Pettersen *et al*, 2021). The model shown in this panel corresponds to the *Saccharomyces cerevisiae* enzyme and is derived from data of Hallett and collaborator (Hallett *et al*, 2022), as deposited in the RCSB database (PDB: 7QCD). (C) Cryo-EM structure of the budding yeast SMC5/6 dimer and NSE5/NSE6 subcomplex interacting with the SMC5/SMC6-head neck region. The lower portion of the panel is derived from the SMC5/6-8mer complex structure of Li and colleagues (Li *et al*, 2023), as deposited in the RCSB database (PDB: 8T8F). Figure created using UCFC ChimeraX and Adobe Illustrator.

In a broad sense, eukaryotic SMC complexes are large DNA binding/modulating ATPases that share several biochemical activities. Cohesin, condensin and the SMC5/6 complex have all been shown to form loops in DNA using an ATP hydrolysis-dependant loop extrusion activity (Kim *et al*, 2023; Davidson & Peters, 2021; Davidson *et al*, 2019; Moon & Ryu, 2023). The mechanism of loop extrusion facilitated by the SMC5/6 complex is similar to the process catalysed by cohesin, which performs two-sided loop-extrusion as a dimeric complex. However, cohesin can also extrude DNA loops as a monomer similar to condensin (Pradhan *et al*, 2023). Thus, in addition to the structural similarities shared by all eukaryotic SMC complexes, DNA loop extrusion is considered one of the unifying features of this family of proteins. How this biochemical activity translates into effective maintenance of chromosome integrity is a topic of active investigation. An excellent discussion of loop extrusion models relevant to SMC complexes and their impact on chromatin structure can be found in several recent reviews (Kim *et al*, 2023; Davidson & Peters, 2021; Davidson *et al*, 2019; Moon & Ryu, 2023; Higashi & Uhlmann, 2022; Barth *et al*, 2023). While it is evident that the SMC5/6 complex shares common features with its sister complexes, it is equally clear that it diverges from other family members in many ways, making it the most enigmatic of the three SMC complexes.

1.4 DISTINCTIVE FEATURES OF THE SMC5/6 COMPLEX

The impact of SMC family proteins on chromosomes is often perceived through the lens of cohesin and condensin's role as key effectors of global changes in chromosome morphology. However, the scope of SMC5/6 complex action is substantially different from that of canonical SMC complexes. The SMC5/6 holoenzyme acts at a more local and targeted level on chromosomes to promote DNA transactions essential for DNA repair and genome stability (Figure 1.2). Until recently, the exact nature of DNA transactions promoted by the SMC5/6 complex was not fully understood, but recent advances on the structural organization, motor activity and biochemistry of the SMC5/6 complex contributed crucial new insights into the specific role of this holoenzyme in DNA repair and replication. We review below distinct features of the SMC5/6 complex that differentiates it from canonical SMC complexes and how these promote the unique repertoire of functions of this holoenzyme (Figure 1.2). For a more detailed review of the common features of SMC complexes, we direct readers to several excellent reviews recently published on this topic (Bürmann & Löwe, 2023; Hoencamp & Rowland, 2023; Kim *et al*, 2023).

The SMC5/6 complex is a post-translational modification enzyme

The SMC5/6 complex distinguishes itself from other SMC complexes by virtue of its unique role as a post-translational modification (PTM) enzyme. The PTM activities of the complex are mediated by two of its subunits, NSE1 and NSE2/MMS21 (Figure 1.1B), that contain RING-like domains known to stimulate ubiquitin and SUMO transfer on substrates, respectively (Potts & Yu, 2005; Zhao & Blobel, 2005; Andrews *et al*, 2005; Doyle *et al*, 2010). The best characterized of these two subunits, NSE2, is a major regulator of multiple functions in eukaryotes, as outlined in

Table 1.1 (Solé-Soler & Torres-Rosell, 2020). Notably, substrates of NSE2-mediated SUMOylation include the SMC5/6 complex itself (Andrews *et al*, 2005; Potts & Yu, 2005; Zhao & Blobel, 2005; Pebernard *et al*, 2008; Albuquerque *et al*, 2013; Bermúdez-López *et al*, 2015; Zapatka *et al*, 2019) along with an impressive number of proteins implicated in genome organisation, DNA replication and DNA repair (Table 1.1).

How are SMC5/6-mediated SUMOylation reactions impacting genome stability? One of the primary functions of the SMC5/6 complex is to limit the formation of toxic recombination intermediates or help resolve them when formed at replication forks. In line with this, self-SUMOylation of the SMC5/6 complex by NSE2, especially in the coiled-coil domain of SMC5, is upregulated at sites of stalled replication forks (Zapatka *et al*, 2019). Also, NSE2-mediated SUMOylation of the yeast RecQ-family helicase SGS1 and its binding partners TOP3 and RMI1 (STR) allows the removal of recombination intermediates at DNA repair sites in a two-step process (Li *et al*, 2022; Agashe *et al*, 2021), specifically: self-SUMOylation of the SMC5/6 complex, an event that promotes STR loading at damaged sites, followed by additional SUMOylation events on SGS1 and TOP3 to increase the efficiency of the recombination function by STR. Similar patterns of SUMOylation are observed with the human Bloom syndrome helicase (BLM), a SGS1 homolog (Bermúdez-López *et al*, 2016; Bonner *et al,* 2016). Interestingly, binding of the SMC5/6 complex to collapsed forks or other damaged structures in chromosomes can promote NSE2-mediated SUMOylation of repair proteins, including Replication protein A (RPA) and RAD59 (Whalen *et al*, 2020), as well as other SMC complex components (McAleenan *et al*, 2012; Wu *et al*, 2012; Albuquerque *et al*, 2013). Altogether, proteins involved in DNA replication, DSB repair,

rDNA maintenance and telomeric length regulation (Figure 1.2) have been shown to be direct targets of the SUMOylation activity of NSE2 (see Table 1.1 for an extensive list and (Räschle *et al*, 2015)).

Table 1. 1: Targets of the SMC5/6 complex SUMOylation, ubiquitylation and interaction reactions

Targets	Biological pathways	References
SUMOylation by NSE2		
SMC5-SMC6	MPH1-mediated fork regression, DNA repair	(Potts & Yu, 2005; Albuquerque *et al*, 2013; Hang *et al*, 2015; Zapatka *et al*, 2019)
NSE2-NSE3-NSE4	DNA damage response pathway	(Andrews *et al*, 2005; Potts & Yu, 2005; Pebernard *et al*, 2008)
Cohesin (SMC1/SMC3/SCC1)	DSB repair/mitosis/rDNA maintenance	(McAleenan *et al*, 2012; Wu *et al*, 2012; Yuen & Gerton, 2018).
Condensin (SMC2/SMC4/BRN1/YCS4/YCG1)	rDNA maintenance	(Albuquerque *et al*, 2013; Yuen & Gerton, 2018)
RPA-RNA pol1 and RAD59	Homologous	(Heideker *et al*, 2011;

	recombination	Whalen *et al*, 2020)
STR (SGS1-TOP3-RMI1), BLM	Double Holliday junction removal	(Li *et al*, 2022; Agashe *et al*, 2021)
POL2, MCM6, RFA1,2,3	DNA replication	(Hang *et al*, 2015; Meng *et al*, 2019; Whalen *et al*, 2020)
TRAX	DNA repair	(Potts & Yu, 2005)
YKU70	DSB repair by non-homologous end joining	(Zhao & Blobel, 2005)
RAD52	Recombinational repair in rDNA	(Torres-Rosell *et al*, 2007)
Shelterin complex	Telomere length maintenance	(Potts & Yu, 2007)
FOB1	rDNA maintainance	(Zapatka *et al*, 2019)
Kinetochore proteins	Recombination at centromere	(Yong-Gonzales *et al*, 2012)
Zebularine induced MET1-DNA protein crosslink	Repair process in Arabidopsis	(Dvořák Tomaštíková *et al*, 2023)
Ubiquitination by NSE1		
NSE3-NSE4	Replication stress	(Kozakova *et al*, 2015;

		Kolesar *et al*, 2022)
RNA polymerase I – RPA190	Ribosomal DNA biosynthesis	(Ibars *et al*, 2023)
RPL24B, CDC73, YHB1, ERG9, CDC48 and POL30/PCNA	Ribosome biosynthesis and metabolism	(Ibars *et al*, 2023)
MMS19	Cytosolic iron-sulfur cluster assembly pathway	(Weon *et al*, 2018)
Protein-protein interactions		
MPH1 helicase	Replication fork remodeling	(Chen *et al*, 2009)
STR (SGS1-TOP3-RMI1)	Resolution of recombination product, Holliday junctions	(Agashe *et al*, 2021)
SRS2, MUS81-MMS4	Recombination intermediate formation and maturation, Holliday junctions	(Agashe *et al*, 2021)
DNA topoisomerase II alpha (TOP2A)	Resolution of DSB-repair intermediates	(Verver *et al*, 2016)

POL2	Replication-S-phase checkpoint	(Winczura *et al*, 2019)
GPS1 (G-protein pathway suppressor 1)	Signaling pathway	(Horváth *et al*, 2020)
Telomeric proteins	Telomere maintenance	(Moradi-Fard *et al*, 2016)
FANCD2-I	Fanconi-anaemia pathway, repair interstrand crosslink	(Rossi *et al*, 2020)
Cohibin and CLIP	rDNA stability	(Moradi-Fard *et al*, 2021)

In contrast to the well-documented SUMO ligase activity of NSE2, much less is known regarding the ubiquitin ligase/E3 activity of NSE1. This protein forms a subcomplex with NSE3-4 proteins and associates with the ATPase head region of the SMC5/6 complex (Sergeant *et al*, 2005; Palecek *et al*, 2006; Duan *et al*, 2009). Interestingly, NSE1 can promote the ubiquitylation of NSE3 and NSE4, but how this impacts the SMC5/6 complex function is still under investigation (Kozakova *et al*, 2015; Kolesar *et al*, 2022). Advanced proteomics analysis has identified Rpa190, Rpl24b, Cdc73, Yhb1, Erg9, Cdc48 and Pol30/PCNA as putative targets for NSE1 for ubiquitination (Ibars *et al*, 2023). This repertoire of substrates suggests roles beyond DNA repair for NSE1-mediated ubiquitination, notably in ribosome biogenesis and metabolism (Table 1.1) (Ibars *et al*, 2023). One of the known functions of NSE1 E3 ligase activity is in RNA polymerase I/RPA190 ubiquitination and subsequent RNA pol I degradation. This process facilitates the SMC5/6-

dependent separation of the ribosomal DNA locus in mitosis (Ibars *et al*, 2023). Beyond this, little is known on the role of NSE1 in substrate ubiquitylation, outlining a clear area for future expansion of research efforts.

The SMC5/6 complex is an unusual nucleic acid binding machine

One of the fundamental activities of all SMC complexes is DNA binding, but the DNA binding preference of the SMC5/6 holoenzyme is quite distinct from that of other SMC complexes. Specifically, components of the SMC5/6 complex demonstrate high affinity for structured DNA like Holliday junctions (HJ) and splayed Y structures, including ssDNA-dsDNA junctions and other types of recombination intermediates (Figure 1.5 Step 1) (Roy *et al*, 2015; Zabrady *et al*, 2016; Alt *et al*, 2017; Serrano *et al*, 2020; Chang *et al*, 2022; Tanasie *et al*, 2022). The higher affinity of the holoenzyme for these substrates (compared to B-form DNA) suggests that the SMC5/6 complex can achieve differential association to chromosomes based on the nature of the DNA lesion/intermediate present at specific positions of the genome. Importantly, the higher abundance of conventional DNA compared to sites of DNA damage *in vivo* must be considered when assessing the impact of modest differences in DNA binding affinities. It is interesting to note that the human SMC5/6 complex shows a clear affinity towards short RNA-DNA hybrid structures that are common transcriptional intermediates known to stall replication fork progression (Serrano *et al*, 2020). In addition, both human and yeast SMC5/6 complexes can bind and stabilize supercoiled/catenated DNA structures in an ATP dependent manner (Serrano *et al*, 2020; Gutierrez-Escribano *et al*, 2020). The binding affinity of the SMC5/6 complex towards unconventional nucleic acid structures is a defining property of this complex, consistent with a

function for the holoenzyme in the repair of DNA lesions formed at specific sites on chromosomes.

The SMC5/6 complex adopts a distinctive architecture

Recent advances in the structural characterization of SMC5/6 complex components provided exciting new insights into the global architecture of the holoenzyme and topology of its subunits (Gutierrez-Escribano *et al*, 2020; Serrano *et al*, 2020; Hallett *et al*, 2021, 2022; Jo *et al*, 2021; Taschner *et al*, 2021; Yu *et al*, 2021, 2022). These studies reveal how the SMC5/6 complex uses distinctive structural features associated with its NSE subunits (Figure 1.1) to perform its unique repertoire of cellular functions.

Unsurprisingly, the SMC5 and SMC6 proteins at the core of the complex share a general architecture similar to that of other SMC family members, including long antiparallel coiled-coil arms connecting the hinges and ATPase head domains of each SMC subunit (Figure 1.1A) (Anderson *et al*, 2002; Haering *et al*, 2002; Serrano *et al*, 2020). Biochemical experiments with purified human and budding yeast SMC5/6 complexes have shown that SMC5 and SMC6 arms align in close proximity to each other throughout most of their lengths, adopting a rod-shaped architecture instead of the open-ring configuration that historically has been attributed to cohesin (Figure 1.1) (Serrano *et al*, 2020; Taschner *et al*, 2021; Yu *et al*, 2021; Hallett *et al*, 2022). More recent small-angle X-ray scattering experiments show that the yeast cohesin complex can adopt –at least transiently– a rod-like structure with juxtaposed coiled-coils, suggesting that this arm configuration is a conserved feature of all SMC complexes (Soh *et al*, 2015). Importantly, the

arm regions of cohesin and condensin are longer than those of the SMC5/6 complex (Bürmann *et al*, 2019; Serrano *et al*, 2020) and bend sharply at so-called pseudo-elbow sites causing the hinge to contact the head-proximal coiled-coil or head-bound non-SMC proteins (Lee *et al*, 2020; Kong *et al*, 2020; Bürmann *et al*, 2019; Davidson *et al*, 2019; Ganji *et al*, 2018; Shaltiel *et al*, 2022). In comparison, the arm region of the SMC5/6 holoenzyme bends slightly (Figure 1.1B) but does not appear to adopt a conventional elbow-bending configuration (Adamus *et al*, 2020; Serrano *et al*, 2020; Jo *et al*, 2021; Yu *et al*, 2021).

The central hinge regions of the SMC5 and SMC6 subunits are also diverged from the corresponding regions of canonical SMCs. X-ray crystal structure of fission yeast SMC5/6 proteins reveals that the hinge region forms a toroidal structure much akin to that of the canonical SMCs. However, the hinge region consists of an unusual 'molecular latch' in SMC5 and a functional 'hub' in SMC6, two conserved interfaces that are absent in cohesin and condensin SMC components. Introduction of mutations in these interfaces affects the functionality of the complex *in vivo* (Alt *et al*, 2017).

Whereas the SMC components of the SMC5/6 complex show interesting differences with other family members, it is clear that the distinctiveness of the SMC5/6 complex arises mostly from the NSE components of the holoenzyme. Early biochemical studies indicated that the NSE elements of the SMC5/6 complex form three subcomplexes; namely NSE1/NSE3/NSE4, NSE2/SMC5, and NSE5/NSE6 (McDonald *et al*, 2003; Pebernard *et al*, 2004, 2006; Sergeant *et al*, 2005; Palecek *et al*, 2006; Duan *et al*, 2009) (Figure 1.1B-C). Human NSE4 has also been found to exist in an

alternate complex with E3 (TRIM31) and MAGE-A1 proteins (Kozakova *et al*, 2015). Overall, the positioning of NSE subcomplexes relative to the core SMC5-SMC6 heterodimer is highly conserved across the eukaryotic kingdom (Bürmann & Löwe, 2023; Hoencamp & Rowland, 2023; Kim *et al*, 2023). The kleisin-family protein NSE4, through its N-terminal helix turn helix (HTH) domain, binds to SMC6 neck region and also connects to SMC5 ATPase head domain using its C-terminal winged helix (WH) region (Figure 1.1B)(Li *et al*, 2023; Palecek *et al*, 2006). Furthermore, kleisin-interacting tandem winged-helix element (KITE) proteins NSE1-NSE3 interact with NSE4 and bridge the entire SMC5-SMC6 head domain area (Figure 1.1B) (Palecek *et al*, 2006; Hallett *et al*, 2022; Yu *et al*, 2022). The NSE2/Mms21 subunit binds to the mid arm region of SMC5 –an unconventional position for NSE binding– where it can promote SUMO transfer on SMC5 (and other substrates; Table 1.1) (Figure 1.1B) using its RING-like domain (Andrews *et al*, 2005; Potts & Yu, 2005; Zhao & Blobel, 2005).

The NSE5-NSE6 dimer, was initially believed to associate with the SMC5/SMC6 heterodimer via the SMC hinge region in early studies on the budding yeast enzyme (Duan *et al*, 2009). Some studies, however, indicate that the NSE5-NSE6 dimer binds to the SMC coiled-coil arm region closer to NSE2 and proximal to the ATPase head of the holoenzyme. The proximity of NSE2 and NSE5/6 on the SMC arms of the complex likely reflects a functional link connecting these proteins since it was shown that NSE5/6 co-ordinates NSE2-mediated SMC5 and SMC6 SUMOylation (Yu *et al*, 2021). The most recent cryo-EM structure of the yeast SMC5/6 complex clearly indicates that the NSE5/6 subcomplex associates with the ATPase head domain and the adjacent coiled-coil neck region of SMC6 (Figure 1.1C). Interestingly, NSE6 competes with NSE4 in its binding to

the neck region of SMC6 and thus regulates the ATPase activity of the complex (Li *et al*, 2023). This situation has additional implications for DNA binding because the kleisin-SMC neck interaction is generally considered a gateway for DNA entry into the ring structure of all SMC complexes. In this context, alternative binding of the kleisin/NSE4 and NSE6 subunits to the SMC neck region of the SMC5/6 complex creates a process of 'gate switching' that is a unique functional feature of this complex (Li *et al*, 2023). Together, the specialized topology and biochemical properties of NSE subunits within the SMC5/6 complex (Figure 1.1) enable the diverse and context-specific roles of this enzyme in the maintenance of genomic integrity. How this functional diversity is achieved is a key question addressed in the following section of the this chapter.

1.5 CELLULAR FUNCTIONS OF THE SMC5/6 COMPLEX: A DNA REPAIR FACTOR WITH A TWIST

Genes encoding components of the SMC5/6 complex have first been identified in genetic screens for mutants affecting the resistance of yeast to DNA damaging agents, consistent with an important role for the complex in the DNA damage response (Nasim & Smith, 1975; Prakash & Prakash, 1977; Lehmann *et al*, 1995). Whereas the involvement of the SMC5/6 complex in DNA repair cannot be overstated, it is equally important to recognize that the cellular roles of the holoenzyme exceed the realm of DNA repair as it also participates in biochemical transactions involving undamaged DNA during interphase and controls effective segregation of chromatin in mitosis. Hence, unlike canonical SMC complexes which have primary roles in the global organization of chromosomes at specific stages of the cell cycle, the SMC5/6 complex acts at a more local level to regulate or otherwise promote various DNA repair and non-repair transactions

throughout the cell cycle (Kegel & Sjögren, 2010; Palecek, 2018) (Figure 1.2). As such, the primary roles of the SMC5/6 complex appear more diverse than those of canonical SMC complexes, which raises the question of whether the SMC5/6 complex acts in those multiple pathways using a unified/single mode of action or by taking advantage of selected activities residing within the complex to achieve functional diversity. We discuss these possibilities in detail in the next two sections of this chapter.

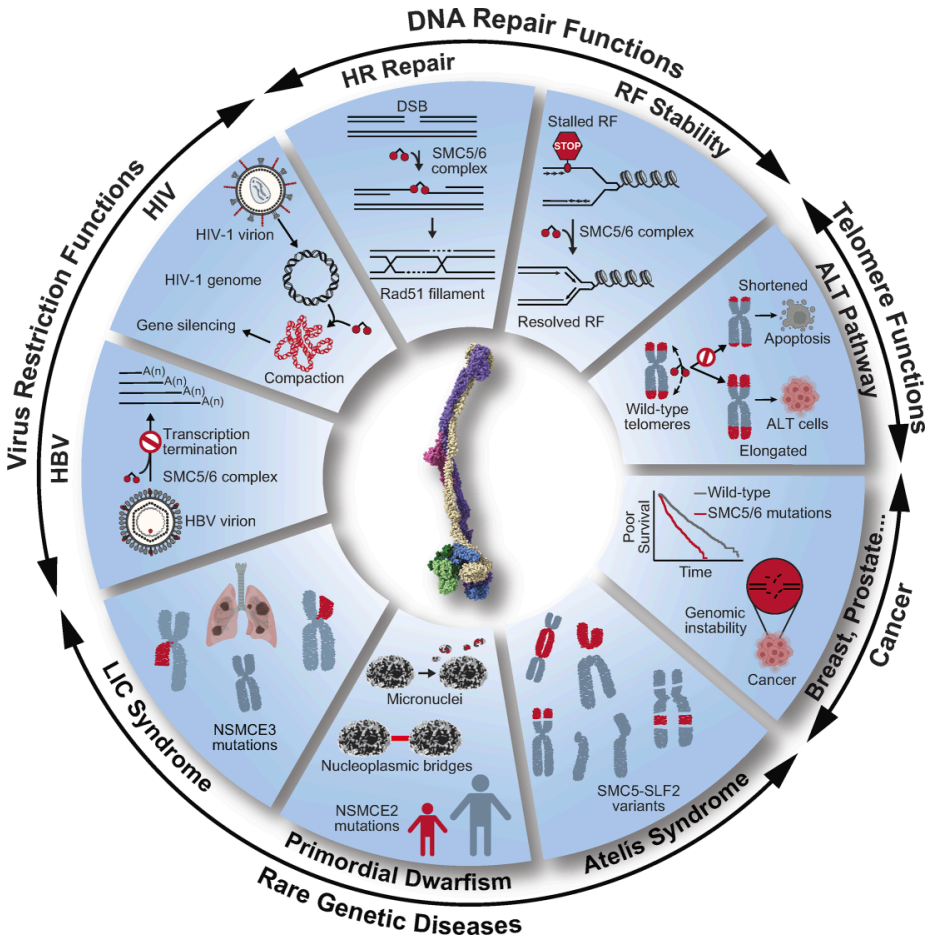


Figure 1. 2: Landscape of the SMC5/6 complex functions and phenotypes. Each section of the figure depicts a distinct cellular function (DNA repair/telomere) and/or disease state associated with the SMC5/6 holoenzyme (viral infection, rare genetic diseases, and cancer). Abbreviations: HR, homologous recombination; RF, replication fork; ALT, alternative lengthening of telomeres; MV, mosaic variegated; LICS, lung disease immunodeficiency and chromosome breakage syndrome; HBV, hepatitis B virus; HIV, human immunodeficiency virus. See text for a detailed

description of the molecular pathways and diseases presented in this figure. Figure created using Biorender and Adobe Illustrator.

Roles in repair of damaged DNA by homologous recombination

One of the most important roles of the SMC5/6 complex *in vivo* is the promotion of DNA double strand break (DSB) repair by homologous recombination (HR; Figure 1.2) (Lehmann *et al*, 1995; Ampatzidou *et al*, 2006; Cost & Cozzarelli, 2006; Potts *et al*, 2006; Kegel & Sjögren, 2010). The SMC5/6 complex facilitates recombination between sister chromatids by recruiting cohesin in the area surrounding damaged DNA, thus allowing it to hold sister chromatids in close proximity. During this process, the SMC5/6 complex SUMOylate the subunits of cohesin and promotes error-free progression of HR repair. In support of this notion, depletion of human NSE2 lead to a decrease in the loading of cohesin to DSB sites (Potts *et al*, 2006). Also, in presence of the defective SMC5/6 complex and cohesin, sister chromatids are misaligned prior to recombination, thereby restricting DNA strand invasion (Potts *et al*, 2006; Ström & Sjögren, 2007; Jeppsson *et al*, 2014; Palecek, 2018). Additionally, the SMC5/6 complex can act in conjunction with proteins like MUS81/MMS4 or STR/BTR (BLM-TOPIIIα-RMI1/2) in stimulating the resolution of complex recombination intermediates generated in the late stages of HR (Figure 1.3). As is the case for cohesin, the SUMO-ligase activity of the SMC5/6 complex plays a vital role in the SUMOylation of several proteins involved in the dissolution of HR intermediates (as discussed earlier; see full list in Table 1.1) (Ampatzidou *et al*, 2006; Bermudez-Lopez *et al*, 2010; Kegel & Sjögren, 2010; Agashe *et al*, 2021).

Roles in the promotion of error-free DNA replication across the genome

Multiple lines of evidence support a conserved role for the SMC5/6 complex in the promotion of proper genome replication, both in the presence and absence of DNA damage (Figure 1.2) and (Figure 1.3). In budding yeast, the SMC5/6 complex binds to chromatin at replication initiation sites in early S phase (Betts Lindroos *et al*, 2006; Kegel *et al*, 2011). The SMC5/6 complex is also directly involved in SUMO-based regulation of multiple important factors in the replisome, one of the most important among them being POL2, the catalytic subunit of DNA polymerase ε crucial for the initiation of DNA replication (Meng *et al*, 2019; Winczura *et al*, 2019). Moreover, SMC5/6 works together with proteins such as STR to facilitate replication completion (Agashe *et al*, 2021).

Interestingly, the SMC5/6 complex exhibits an important role in the maintenance of proper chromosome topology during DNA replication which allows faithful completion of replication and prevents chromosome fragmentation during mitosis. Indeed, the separation of the two parental strands during DNA replication creates positively supercoiled DNA ahead of advancing replication forks, a situation that can hinder replication fork progression if not addressed. The main players involved in the removal of replication-induced topological stress are topoisomerases TOP1 and TOP2 (Bermejo *et al*, 2007) along with the SMC5/6 complex itself (Jeppsson *et al*, 2014). In yeast, the SMC5/6 complex associates with chromosome during S-phase in a length-dependant manner to decrease topological stress during replication especially when TOP2 activity is impaired (Jeppsson *et al*, 2014). The impact of positive supercoils formed at replication forks can also be mitigated by rotating replication forks, which creates sister chromatids intertwinements (SCI) or catenation behind advancing forks. The SMC5/6 complex sequesters these SCIs behind the replication fork by facilitating fork rotation, a process that ultimately promotes normal

replication progression (Postow *et al*, 2001; Wang, 2002; Kegel & Sjögren, 2010; Kegel *et al*, 2011). Hence, these observations point out to a conserved prerequisite for the SMC5/6 complex in successful completion of genome replication.

The SMC5/6 complex also promotes stabilization of replication forks at sites of DNA lesion to prevent fork collapse, mainly by limiting the formation and/or helping in the resolution of recombination intermediates at stalled replication forks (Irmisch *et al*, 2009) and (Figure 1.3). The SMC5/6 complex maintains the integrity of stalled replication forks by loading and SUMOylating RPA and RAD52 proteins, thus keeping forks in a recombination-competent configuration (Irmisch *et al*, 2009; Whalen *et al*, 2020). Advancing replication fork can also bypass a DNA lesion in a HR-dependant template switch mechanism where SMC5/6 complex along with STR can participate in the resolution of recombination intermediates formed after the template switch (Branzei *et al*, 2008; Sollier *et al*, 2009; Choi *et al*, 2010; Menolfi *et al*, 2015).

In yeast, the SMC5/6 complex physically interacts with the MPH1 helicase, a homolog of Fanconi anemia protein M, in a manner that prevents replication fork reversal by MPH1 (Figure 1.3)(Chen *et al*, 2009; Lafuente-Barquero *et al*, 2017). Genetic data in fission yeast also suggest that the NSE5/6 subunits, as a part of the SMC5/6 complex, play a crucial role in preventing the accumulation of Holliday junctions when replication forks encounter DNA lesions (Pebernard *et al*, 2006). Furthermore, mutagenesis data from a recent study reveals that the NSE5/6-SMC6 interaction promotes resolution of DNA repair intermediates but not replication termination (Li *et al*, 2023). In higher eukaryotes, the SMC5/6 complex has been shown to associate with the

FANCD2 and FANCI proteins of the Fanconi Anaemia pathway to promote recombination repair at stalled replication forks (Rossi *et al*, 2020). Consistent with the observation that the SMC5/6 complex plays crucial roles during DNA replication, SMC5 depletion in mouse cells resulted in increased replication fork stalling that was associated with higher apoptosis in central cortical tissues along with reduced cortex sizes (Atkins *et al*, 2020).

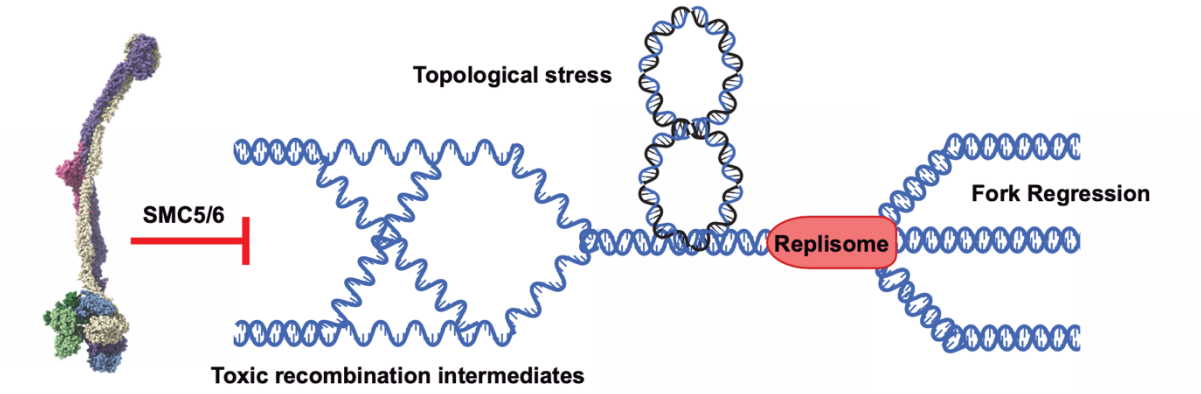


Figure 1. 3: The schematic depicting the role of the SMC5/6 complex (on left) in topological management of the genomic DNA including the resolution of toxic recombination intermediates, topological stress created at replication forks and replication fork regression. Figure created using Adobe Illustrator.

Promoting replication at hard-to-replicate genomic loci

Natural sites on chromosomes that act as replication barriers or that are otherwise difficult to replicate represent a unique type of challenge for genome stability, one whose effective resolution is especially dependant on SMC5/6 complex activity. This situation is well exemplified by the role of the SMC5/6 complex at replication fork barrier (RFB) sites in the ribosomal DNA (rDNA) array of yeast (Pebernard *et al*, 2008; Peng *et al*, 2018; Moradi-Fard *et al*, 2021).

Chromatin immunoprecipitation studies have shown that the SMC5/6 complex associates with chromosomal DNA across the genome, but is particularly enriched at the rDNA array on budding yeast chromosome XII during S phase (Torres-Rosell *et al*, 2005; Ampatzidou *et al*, 2006; Betts Lindroos *et al*, 2006). Because of its highly repetitive and heterochromatic nature, the rDNA array often experiences a higher frequency of defective replication relative to other regions of the genome. This is compounded by the unidirectional nature of DNA replication at this locus, which prevents the rescue of stalled replication forks by forks converging from the other side of the chromosome. As expected, rDNA replication delay and instability are prevalent in *smc5/6* mutants, reflecting the special contribution of the SMC5/6 complex to the maintenance of rDNA stability. Importantly, defects in rDNA replication in *smc5/6* mutants can cause subsequent chromosome non-disjunction and a higher frequency of Holliday junctions, thus impairing chromosomal segregation in mitosis (Torres-Rosell *et al*, 2005, 2007; Murray & Carr, 2008).

Roles as a regulator of the DNA transactions in heterochromatin

Apart from the rDNA locus discussed above, localization analyses of the SMC5/6 complex have revealed that the complex is enriched within heterochromatic-like regions at telomeres (Pebernard *et al*, 2008; Moradi-Fard *et al*, 2016; Moradi-Fard *et al*, 2021). This enrichment appears to be functionally important for the maintenance of telomere length (Figure 1.2). In telomerase negative cancer cells, the NSE2 subunit SUMOylates multiple telomere binding proteins to promote telomeric HR and lengthening by alternative lengthening of telomeres (ALT) (Figure 1.2; Table 1.1) (Potts & Yu, 2007). In budding yeast, loss of the SMC5/6 complex activity in cells lacking telomerase enhances senescence via a HR independent pathway (Noël &

Wellinger, 2011). Together, these studies suggest that the maintenance of telomeres by the SMC5/6 complex is evolutionary conserved. Furthermore, the SMC5/6 complex stabilizes telomeres through its involvement in the processing of recombination intermediates at chromosome ends (Pebernard *et al*, 2008; Moradi-Fard *et al*, 2016).

Repair of DSB within heterochromatic regions of the genome is surprisingly different from that of DSBs in euchromatin. Late events of HR involving RAD51-mediated strand invasion occurs usually outside of the heterochromatin domain in Drosophila (Chiolo *et al*, 2011). Similarly in yeast, HR of DSBs in the heterochromatin-like rDNA requires re-localization of damaged DNA to extranucleolar sites before it can be repaired (Torres-Rosell *et al*, 2007; Horigome *et al*, 2016). HR is supressed at the rDNA locus through the SUMOylation of RAD52 by the SMC5/6 complex in yeast, an event that prevents RAD52 foci formation in the nucleolus (Torres-Rosell *et al*, 2007). Importantly, this event supresses recombinational loss of rDNA repeats on chromosome XII, thus ensuring copy number stability at this locus (Torres-Rosell *et al*, 2007). Likewise, in metazoans like flies and mice, the SMC5/6 complex and its SUMO-ligase activity are essential for recruiting and co-ordinating nuclear actin and myosin proteins to relocate heterochromatic DSBs to the nuclear periphery before strand invasion (Ryu *et al*, 2015; Caridi *et al*, 2018). This phenomenon allows error-free progression of HR repair within heterochromatin and limits ectopic recombination between repeated regions of the genome (Ryu *et al*, 2015; Caridi *et al*, 2018; Palecek, 2018).

Despite substantial progress in recent years, the mechanistic basis for SMC5/6 complex-

dependent movement of DNA damage across specific nuclear compartments is not fully understood. It is interesting to note that SMC5 and SMC6 both show significant affinity for microtubules (Laflamme *et al*, 2014) and microtubules can promote non-linear/directional motion of damaged chromatin in cells (Oshidari *et al*, 2018). It is therefore tempting to speculate that the SMC5/6 complex might act as a bridge connecting DNA damage to microtubules during DNA repair reactions. More work is required to shed light on this exciting, but putative mode of action.

Roles as a master silencer of episomal and viral DNA

Silencing of episomal DNA templates is an unexpected function specifically attributed to the SMC5/6 holoenzyme and does not appear to extend to other eukaryotic SMC complexes (Figure 1.2). Interestingly, the bacterial Wadjet defence system —where three Wadjet subunits form an SMC-family complex— recognizes and protects bacterial hosts from external plasmid transformation in an ATP hydrolysis dependant manner (Deep *et al*, 2022). Multiple recent studies have reported that the SMC5/6 complex can epigenetically silence extrachromosomal viral DNA, helping to defend cells against infections from hepatitis B virus (HBV), human immunodeficiency virus (HIV), herpes simplex virus and papillomavirus (Decorsière *et al*, 2016; Bentley *et al*, 2018; Gibson & Androphy, 2020; Han *et al*, 2022; Irwan *et al*, 2022).

The SMC5/6 complex restricts HBV by entrapment of the viral episome DNA inside the complex (Figure 1.2). This is followed by recruitment of the episomal DNA-protein complex to promyelocytic leukemia nuclear bodies (PML-NB) for subsequent transcriptional repression of

viral DNA by NSE2 (Oravcová *et al*, 2022; Abdul *et al*, 2022). The recruitment step depends on SIMC1, a paralog of SLF1 that forms a NSE5/NSE6-like complex with SLF2 and behave similarly to its yeast counterpart in the localization of the SMC5/6 complex to specific regions of chromosomes (Figure 1.4). The virus silencing function of the SMC5/6 complex is not limited to HBV since it can also silence HIV1 DNA provirus before integration into the host cell chromosomes by inhibitory epigenetic modification (Figure 1.2). This process is mediated mainly by NSE2-dependent SUMOylation (Dupont *et al*, 2021; Irwan *et al*, 2022). Furthermore, the SMC5/6 complex appears to act as a global guardian against virus infection because several recent studies reported similar virus-restriction roles of the SMC5/6 complex in epigenetic silencing and/or inhibition of DNA replication of Epstein–Barr virus (EBV) (Yiu *et al*, 2022), Kaposi's sarcoma herpesvirus (KSHV) (Han *et al*, 2022), human papilloma virus (HPV) (Gibson & Androphy, 2020), and herpes simplex virus 1 (Xu *et al*, 2018).

The role of the SMC5/6 complex in episome silencing appear to exceed the domain of viral genomes since a recent study has revealed a potential role for the budding yeast SMC5/6 complex –specifically the NSE2 subunit– in restricting the parasitic 2µ plasmid (Hays *et al*, 2020). It will be interesting for future studies to define the full repertoire of episomes targeted by the SMC5/6 complex *in vivo* and whether it achieves silencing of extrachromosomal DNA by a universal mechanism applicable to all episome types across the eukaryotic kingdom.

1.6 THE MODE OF ACTION OF THE SMC5/6 COMPLEX: RECOGNITION OF DNA SUBSTRATES AND SUBSEQUENT ACTIVATION

The involvement of the SMC5/6 complex in wide range of cellular functions implies its activity must be precisely controlled in different chromosomal contexts. How context-specific activation of the SMC5/6 complex is achieved is unclear, but the consequences of allowing unrestricted activation of the SMC5/6 complex are likely to be very toxic, highlighting the crucial importance of this issue for cell homeostasis. We discuss in the next section how the SMC5/6 complex precisely and dynamically recognize its substrates *in vivo* and what impact it has on chromosomal DNA once the holoenzyme is activated.

Basal DNA binding activity of the SMC5/6 complex: Who does what?

The SMC5/6 complex associates with DNA via at least three general types of intrinsic DNA binding activities which are not mutually exclusive and often rely on each other for the execution of specific cellular function. The different modes of binding include topological DNA entrapment by the SMC5/6-NSE1/3/4 ring (Gutierrez-Escribano *et al*, 2020; Kanno *et al*, 2015; Taschner & Gruber, 2023), conventional/direct (usually electrostatic) DNA binding, and a specific form of DNA loop entrapment (usually called pseudo-topological entrapment) mediated by selective subunits/domains within the complex (Gutierrez-Escribano *et al*, 2020; Roy & D'Amours, 2011; Roy *et al*, 2011, 2015; (Gutierrez-Escribano *et al*, 2020; Kanno *et al*, 2015; Taschner & Gruber, 2023). The DNA binding domains of the SMC5/6 complex can mediate direct interaction with diverse types of DNA repair and recombination intermediates as evidenced from biochemical analyses (Roy & D'Amours, 2011; Roy *et al*, 2011, 2015; Zabrady *et al*, 2016) as well as from *in*

vivo experiments (Zabrady *et al*, 2016; Etheridge *et al*, 2021).

The core components of the SMC5/6 complex, the SMC5-SMC6 dimer, has its independent DNA binding properties apart from the holocomplex. For instance, structural and biochemical assays show that the hSMC5-SMC6 heterodimer has higher binding affinity towards ssDNA compared to dsDNA (Serrano *et al*, 2020). In-depth DNA binding assays with a series of functional domain fragments revealed that SMC5 and SMC6 monomers can interact with their DNA substrates via two distinct DNA-binding domains on each SMC molecule, one located in the hinge domain and adjacent coiled-coil sequences, and the second one in the ATPase head domain of the protein (Roy *et al*, 2015; Alt *et al*, 2017). Interestingly, formation of the SMC5-SMC6 heterodimer (Figure 1.1A) improves affinity for dsDNA substrates relative to that observed with either of the SMC components in their monomeric forms (Roy *et al*, 2015). Overall, these observations indicate that the basal DNA binding activity of the SMC5/6 complex comes at least in part from the intrinsic ability of SMC5 and SMC6 monomers to associate with nucleic acid substates (Roy & D'Amours, 2011; Roy *et al*, 2011).

Association of the SMC5/6 holoenzyme to chromosomal DNA is also regulated by the NSE subunits of the complex. Additionally, the NSE1/NSE3/NSE4 trimer (Figure 1.1B) demonstrates intrinsic DNA binding activity with higher affinity towards short dsDNA than ssDNA, an activity mainly attributed to the NSE3 subunit (Zabrady *et al*, 2016). The cryo-EM structure of DNA-bound yeast Smc5/6 complex reveals that the NSE1/NSE3/NSE4 subcomplex connects the SMC5/6 ATPase-head domain and locks the SMC5/6 complex encircling the DNA double helix by forming

a clamp (Yu *et al*, 2022). Interestingly, dsDNA entrapment into the ring structure made of SMC5/6-NSE1/3/4 (*i.e*., clamp formation) requires ATP hydrolysis, pointing towards a contribution of the ATPase head domains of SMC subunits in this process (Zabrady *et al*, 2016; Adamus *et al*, 2020; Vondrova *et al*, 2020; Etheridge *et al*, 2021; Jo *et al*, 2021; Chang *et al*, 2022; Hallett *et al*, 2022). During DNA clamp formation, the complex configuration created when the SMC5-SMC6 heterodimer associates with NSE3-NSE4 gives rise to a positively charged central tunnel that stabilizes the negatively charged DNA backbone (Yu *et al*, 2022). Using cryo-EM approach, it was possible to generate a detailed view of the critical DNA binding patches/residues in each of the complex subunit that guide DNA binding and/or clamp formation (Yu *et al*, 2022). Consistent with this view, alterations of individual domains/subunits are met with varied effects on growth and cellular response to DNA damage (Yu *et al*, 2022).

The NSE5-NSE6 subcomplex is another important contributor to the DNA binding activity of the SMC5/6 holoenzyme. Recent studies have revealed that the use of alternative subunits as binding partners for the SMC6 neck region (*i.e*., NSE4 or NSE6) can potentially gate or otherwise influence DNA entry into the SMC5/6 complex (Li *et al*, 2023). The switch from NSE4 to NSE6 as binding partner for SMC6 neck region also modulates the engagement of the SMC6 head domain with that of SMC5 head domain, ultimately regulating the ATPase activity of the complex (Li *et al*, 2023).

Overall, the ATP hydrolysis cycle involves three main structural conformation of the SMC5/6 complex. First, the 'ATP-engaged state' that includes two ATP molecules 'sandwiched' between

the residues of the SMC5-SMC6 head domains maintains adjacent SMC arms at a distance, thus opening the SMC compartment. Second, the 'juxtaposed state' where the heads are disengaged promotes co-alignment of SMC5 and SMC6 arms, yielding a closed SMC compartment. And finally, the 'inhibited state' where NSE5/6 interacts with the arm and heads of the SMC5/6 complex prevents head engagement and ATP hydrolysis (Gutierrez-Escribano *et al*, 2020; Kanno *et al*, 2015; Taschner & Gruber, 2023). This inhibition is counteracted by the binding of a suitable DNA substrate, an event that promotes more stable DNA clamping by the SMC5/6 complex (Yu *et al*, 2022). Overall, the NSE5/6 complex interacts with the SMC5/6 head and arm domains through both stable and dynamic interactions to allow intricate regulation of ATP hydrolysis, DNA segment capture and DNA entrapment/clamping.

Apart from the above-mentioned configurations, another model proposes a pseudo-topological entrapment of DNA in two SMC sub-compartments The SMC5-SMC6 head engagement creates an SMC ('S') and a kleisin ('K') compartment (Taschner *et al*, 2021; Taschner & Gruber, 2023). Additionally, cryo-EM structure revealed that the NSE3 subunit may link the SMC5 and the SMC6 coiled-coil to separate the S compartment into two halves, the upper and the lower SMC compartments (Yu *et al*, 2022). Most up to date cystine crosslinking experiments show that a DNA segment is often effectively held as a loop with (at least) one dsDNA passing through the lower S compartment and another one (or more) through the upper S compartment. This mode of binding is referred to as DNA segment capture and the conformation taken by the SMC5/6 complex during this capture is an important intermediate that connects initial DNA loading to subsequent biochemical activities of the SMC5/6 complex (Li *et al*, 2023; Taschner & Gruber,

2023).

The contribution of several subunits of the SMC5/6 complex to its DNA binding activity combined with the intricate pattern of association of these subunits to chromosomal DNA have important implications for the mechanism of action of the enzyme. During DNA loop extrusion, the multiple DNA binding sites within the SMC5/6 complex might trigger the reeling of DNA while simultaneously holding onto both arms of the forming loop. Also, SMC complexes do not necessarily entrap DNA topologically during the extrusion process (Li *et al*, 2023). It is therefore possible that –in the execution of different cellular functions– the SMC5/6 complex might use different modes of DNA interaction to create DNA loops in the genome (Li *et al*, 2023; Taschner & Gruber, 2023). Indeed, the array of different DNA binding sites within the complex and conformational changes associated with DNA binding are likely to support different loop extrusion models, as observed with cohesin and condensin (Davidson & Peters, 2021). Taken together, these observations indicate that the SMC5/6 complex uses multiple distinct DNA binding domains/modes to affect functional changes in chromatin/DNA configuration.

Context matters: When and how the SMC5/6 complex binds to chromosomes in vivo?

As previously discussed, binding of the SMC5/6 complex to damaged DNA does not appear to impact chromatin at the level of entire chromosomes because genome-level condensation or cohesion are not observed as a direct consequence of SMC5/6 holoenzyme activation. How the complex achieves selectivity in DNA substrate binding and modification *in vivo* is a highly relevant issue for the mode of action of the SMC5/6 holoenzyme. This is accomplished through context-

specific activation during DNA transactions combined with a mechanism of proactive inhibition of the complex in non-targeted regions of the genome.

We now know that several proteins play a concerted role to facilitate the loading of the SMC5/6 complex at specific sites of repair during the DNA damage response. For example, fission yeast BRCT domain-containing protein BRC1, and its budding yeast homologue RTT107 are important recruiters of the complex to chromosomes. Non-SMC elements such as NSE5 and NSE6 also perform an important role in the loading of the SMC5/6 complex on chromatin *in vivo* (Leung *et al*, 2011; Oravcová *et al*, 2019; Etheridge *et al*, 2021; Hallett *et al*, 2021; Taschner *et al*, 2021; Yu *et al*, 2021)(Figure 1.4). In yeast, BRC1/RTT107 recruits the SMC5/6 complex to lesioned/γH2A-covered DNA via the NSE5-NSE6 sub-complex (Leung *et al*, 2011; Oravcová & Boddy, 2019; Etheridge *et al*, 2021) (Figure 1.4). Similarly, human SLF1 (NSE5-like) and SLF2 (NSE6-like) act as a SMC5/6 complex loader through RAD18-mediated binding to damaged DNA sites (Räschle *et al*, 2015). In budding yeast, the NSE5/6 complex also promote loading of the SMC5/6 complex at stalled replication forks (Bustard *et al*, 2012) (Figure 1.4). Interestingly, during viral infection a paralog of SLF1, SIMC1, forms a complex with SLF2 to localize the SMC5/6 complex to virus DNA to restrict virus DNA replication (Oravcová *et al*, 2022), although it is not clear yet whether SIMC1 has roles in other cellular contexts. The NSE5/6-complex is not only essential for directly loading the complex at sites of DNA damage, but also for the stabilization of its binding to dsDNA when the complex bind nucleic acids initially by itself (Etheridge *et al*, 2021) (Figure 1.4).

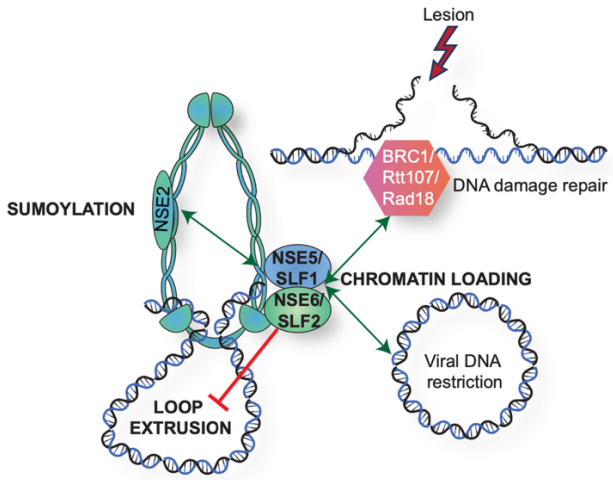


Figure 1. 4: Schematic of NSE5/6 complex functions. The NSE5/SLF1-NSE6/SLF2 complex localizes the SMC5/6 complex to viral replication centers and to DNA lesions via mediator proteins at sites of DNA damage. Other functions of the NSE5/6 complex include the inhibition of loop extrusion mechanism of the SMC5/6 complex and promoting the SUMOylation activity of NSE2. Figure created using Adobe Illustrator.

Chromatin marks such as histone modifications can also act as a cue for the recruitment of the SMC5/6 complex to specific chromosomal regions. For instance, the recruitment of BRCA1 —an important HR promoting factor— requires the recognition of histone H4 unmethylated at lysine 20 (K20me0) through the ankyrin repeat domain of BRCA1-associated RING domain protein 1 (BARD1). The SLF1/SLF2 complex can also specifically recognize and bind histone H4-K20me0 (Nakamura *et al*, 2019). Hence, this provides an intriguing putative mechanism for the selective recruitment of the SMC5/6 complex to newly replicated DNA, an event that may be important for the promotion of HR. Consistent with this observation, the ankyrin repeat domain (ARD) responsible for histone H4-K20me0 recognition of BARD1 is conserved in SLF1 and several other proteins involved in DNA repair and replication, such as TONSL (Tonsoku-like protein) (Nakamura *et al*, 2019). Whether this ARD-dependent recruitment mechanism is operational across the

genome for the SMC5/6 complex is unclear. Interestingly, it has been established that enrichment of the SMC5/6 complex at undamaged regions of the genome, such as centromere and intergenic regions, requires the cohesin loading factor SCC2, which indicates that alternative modes of recruitment likely operate in the absence of DNA damage (Betts Lindroos *et al*, 2006).

Once the SMC5/6 complex is recruited to genomic DNA by chromatin loaders, the first level of regulation of DNA-binding occurs at the level of ATP binding and hydrolysis by the holoenzyme. The SMC5/6 complex interacts with ssDNA in an ATP-independent manner and binding to dsDNA occurs through predominantly ATP-dependant mechanisms especially during clamp formation (Kanno *et al*, 2015; Tanasie *et al*, 2022; Yu *et al*, 2022). The ATP-dependant remodelling of the SMC5/6 complex and subsequent binding to DNA also promotes the SUMO-E3 ligase activity of NSE2 (Varejão *et al*, 2018; Bermudez-Lopez *et al*, 2010). The NSE2 subunit lacks DNA binding domains, but its DNA repair functions require stable docking onto the SMC5 protein (Varejão *et al*, 2018). Specifically, an electrostatic interaction between DNA and a positively charged patch in the ARM domain of SMC5 activates the E3-SUMO-ligase activity of NSE2, which further impacts the loading of the complex on chromatin by self-SUMOylation of the core SMC proteins (Varejão *et al*, 2018; Yu *et al*, 2021). Furthermore, NSE5/6-mediated loading of the SMC5/6 complex on DNA can promote the SUMO-ligase activity of NSE2 (Figure 1.4) (Oravcová & Boddy, 2019; Yu *et al*, 2021).

Recent evidence suggests the existence of additional layers of functional interplay connecting the various subunits and substrates of the SMC5/6 complex. For instance, we now know that the

proximity of NSE2 (in addition to NSE5 and NSE6 subunits) to the ATPase head domain of the SMC components (Yu *et al*, 2021; Li *et al*, 2023; Taschner & Gruber, 2023) can potentially regulate DNA entry into the ring structure of the complex (Pradhan *et al*, 2023) (Figure 1.4). Also interestingly, the NSE5/6 subcomplex compete with high amounts of dsDNA structures to negatively modulate the binding efficiency of the complex to DNA duplex structures (Hallett *et al*, 2021), thus showing both negative and positive impacts on the DNA binding activity of the SMC5/6 complex. This dual impact of NSE5/6 may provide a regulatory mechanism to limit SMC5/6 complex activity to specific regions of chromosomes and allows the complex to adapt to diverse requirements of its ATPase activity during DNA replication and repair actions.

What happens after the SMC5/6 complex binds its DNA substrates in vivo?

Recent biochemical and biophysical advances on the mechanism of the SMC5/6 complex suggest a multi-step catalytic cycle on DNA substrates. One of the earliest steps likely involves SMC5/6 complex translocation along dsDNA by one-dimensional diffusion either as monomeric or oligomeric complex (Tanasie *et al*, 2022; Pradhan *et al*, 2023). Its dynamic interaction with chromatin increases scanning for probable lesions on DNA molecules, especially enriched at ssDNA-dsDNA interfaces (Gutierrez-Escribano *et al*, 2020; Serrano *et al*, 2020; Chang *et al*, 2022; Tanasie *et al*, 2022). Sliding on dsDNA may not be a pre-requisite for binding to ssDNA-dsDNA junction site as the complex can also bind this type of DNA structures directly (Figure 1.5) (Chang *et al*, 2022).

Translocation on DNA is subsequently accompanied by local DNA compaction at sites of DNA lesion, likely involving dimerization of the holoenzyme (Tanasie *et al*, 2022) (Figure 1.5). DNA

compaction by the SMC5/6 complex uses the energy of ATP hydrolysis and proceeds through loop extrusion of chromatin (Pradhan *et al*, 2023) (Figure 1.5), as seen with other SMC proteins. The complex performs symmetrical DNA loop extrusion in a dimer configuration very similar to that of cohesin. The loop-extrusion based DNA compaction activity of the SMC5/6 complex is consistent with its originally assigned role as an intermolecular DNA linker. This is because during compaction, the SMC5/6 complex can link and load multiple molecules of DNA in trans while performing its cellular functions (Tanasie *et al*, 2022; Kanno *et al*, 2015). Recently, the loop extrusion activity of SMC5/6 complex has also been observed during transcription. Single-molecule imaging and Hi-C analyses show that SMC5/6 complex dimers can recognize and link chromosomal regions containing positively supercoiled DNA especially during transcription, and initiate loop extrusion to fold the supercoiled DNA into large plectonemic loops for three-dimensional organization of chromosomes (Jeppsson *et al*, 2024; Diman *et al*, 2023).

As the SMC5/6 complex engages in loop extrusion, its different subunits play distinct and sometimes antagonistic contributions to the overall activity of the enzyme. For instance, the NSE2 subunit is essential for the loop extrusion activity of the SMC5/6 complex, which is a unique property of this enzyme as other SMC complexes do not require an additional subunit to be bound to the coiled-coil arms of SMC proteins to allow loop extrusion. In contrast, the NSE5/6 subunits may act as negative regulators of loop extrusion by the SMC5/6 complex since they seem to reduce loop initiation rate and loop persistence (Figure 1.4). The NSE5/6 subcomplex can also prevent the dimerization of the holoenzyme, which is essential for loop extrusion (Pradhan *et al*, 2023). Interestingly, the NSE5/6 subcomplex does not affect loop extrusion

dynamics after loop initiation has occurred. At first glance, NSE5/6 inhibiting loop extrusion might seem conflicting with the fact that this subcomplex is also a primary loader of the SMC5/6 holoenzyme on chromatin, a precondition for topological loading of DNA into the complex (Taschner *et al*, 2021; Taschner & Gruber, 2023). However, SMC5/6 complex-mediated loop extrusion is known to occur independently of topological entrapment of DNA. The SMC5/6 complex can extrude loops in the absence of NSE5/6 but cannot topologically bind to DNA without NSE5/6 (Pradhan *et al*, 2023).

Overall, the catalytic cycle of the SMC5/6 complex will result in local DNA/chromatin compaction and looping at specific sites of the genome. We hypothesize that this localised compaction of chromatin will create a phenomenon of steric hindrance that diminishes DNA mobility in the vicinity of repair sites, likely shielding fragile DNA intermediates from the action of undesirable modifying enzymes and acting as a cue for the loading of specific DNA repair enzymes (Figure 1.5). Such local restriction of DNA mobility is expected to be transient and not extend beyond the initial recruitment of early DNA repair effectors, consistent with the *in vivo* observed decrease in the mobility of newly-formed DNA lesions followed by a distinct increase at later stages of the DNA repair process (Smith *et al*, 2018, 2019; Hauer & Gasser, 2017; García Fernández & Fabre, 2022). Finally, we anticipate that DNA looping by the SMC5/6 complex will have other impacts on chromatin dynamics. For instance, based on what is observed for cohesin, SMC5/6 complex-mediated chromosome looping might facilitate DNA repair by controlled distribution of the histone phosphorylation mark γH2AX (Arnould *et al*, 2021).

Completing the catalytic cycle: Releasing the SMC5/6 complex from compacted DNA

Similar to the loading of the SMC5/6 complex on damaged DNA, its effective unloading from densely compacted chromatin is critically important to enable effective turnover of the holoenzyme *in vivo* (Figure 1.5), as previously observed with condensin and other chromatin-binding enzymes (Thattikota *et al*, 2018; Chen *et al*, 2004; Cheutin *et al*, 2003; Martin & Cardoso, 2010). Previous studies suggest that the unloading of the SMC5/6 complex in late mitosis is important for effective chromosome segregation since misregulation of the SMC5/6 complex can lead to the formation of anaphase bridges (Gallego-Paez *et al*, 2014; Gomez *et al*, 2013). Despite extensive knowledge of the mechanisms underpinning SMC5/6 complex loading onto chromatids, very little is known about the mechanism responsible for unloading the holoenzyme from chromosomal substrates once its catalytic cycle is completed.

What mechanism might be responsible for SMC5/6 complex release from compacted chromatin? The unloading of the SMC5/6 complex might be similar to its sister SMC complexes since they share some structural and functional features. In the case of condensin, a chromatin rescue factor known as the CDC48 segregase (together with its ubiquitin-adaptor complex UFD1-NPL4) is recruited onto condensin and extracts the trapped enzyme from its compacted product (Thattikota *et al*, 2018). Whether the Cdc48 segregase or a similar enzyme facilitates unloading of the SMC5/6 complex from compacted chromatin remains to be established. However, it is noteworthy that CDC48 has been identified as a target for ubiquitination by the NSE1 subunit of the SMC5/6 complex (Ibars *et al*, 2023), suggesting a likely functional link between these two proteins.

More direct/intrinsic mechanisms for the release of the SMC5/6 complex from chromatin may be involved in its catalytic cycle. For instance, it is possible that the ATPase activities of the complex might generate sufficient mechanical force to promote its release from compacted chromatin (Vondrova *et al*, 2020). Finally, it is conceivable that compacted DNA itself might trigger the dissociation of some subunits from the SMC5/6 complex, thus enabling a reduction in its affinity for chromosomal DNA or otherwise opening the ring structure necessary for topological entrapment (Serrano *et al*, 2020; Roy *et al*, 2015). Consistent with this view, analysis of nuclear extracts by size exclusion chromatography reveals that the interaction connecting SMC5 and SMC6 is weakened or lost specifically during mitosis in human cells (Behlke-Steinert *et al*, 2009), suggesting a possible disassembly of the complex under conditions of generalized chromatin compaction. While the unloading of cohesin and condensin from genomic DNA is often regulated by post-translational modification (PTM), it is important to recognize that non-PTM mechanisms can also promote chromatin unloading of these complexes (Thattikota *et al*, 2018; Murayama & Uhlmann, 2015). Similar mechanisms might also contribute to the release of the SMC5/6 complex from its genomic substrates.

Taken together, extensive progress in deciphering the structure and function of the SMC5/6 complex suggests a 5-step model to explain the catalytic cycle of the holoenzyme on DNA and how it stimulates DNA repair. Chronologically, the first step involves a basal association of the SMC5/6 complex to undamaged DNA using both ATP-dependent and independent mechanisms (step 1; Figure 1.5). This is followed by one-dimensional translocation of the complex on dsDNA,

which leads to dimerization of the holoenzyme at sites of DNA damage (step 2; Figure 1.5). Subsequently, the complex would engage in DNA loop extrusion leading to chromatin compaction (step 3; Figure 1.5). We propose that chromatin compaction at sites of DNA damage will stabilize and protect fragile DNA repair intermediates. Either concomitantly or subsequently with the previous step, the SMC5/6 holoenzyme would recruit and activate DNA repair factors at sites of DNA damage through its post-translational modification activities, thereby promoting downstream steps in the DNA repair process (step 4; Figure 1.5). Completion of repair would finally lead to chromosome unloading of the SMC5/6 holoenzyme, which would then be free to initiate a new catalytic cycle (step 5; Figure 1.5).

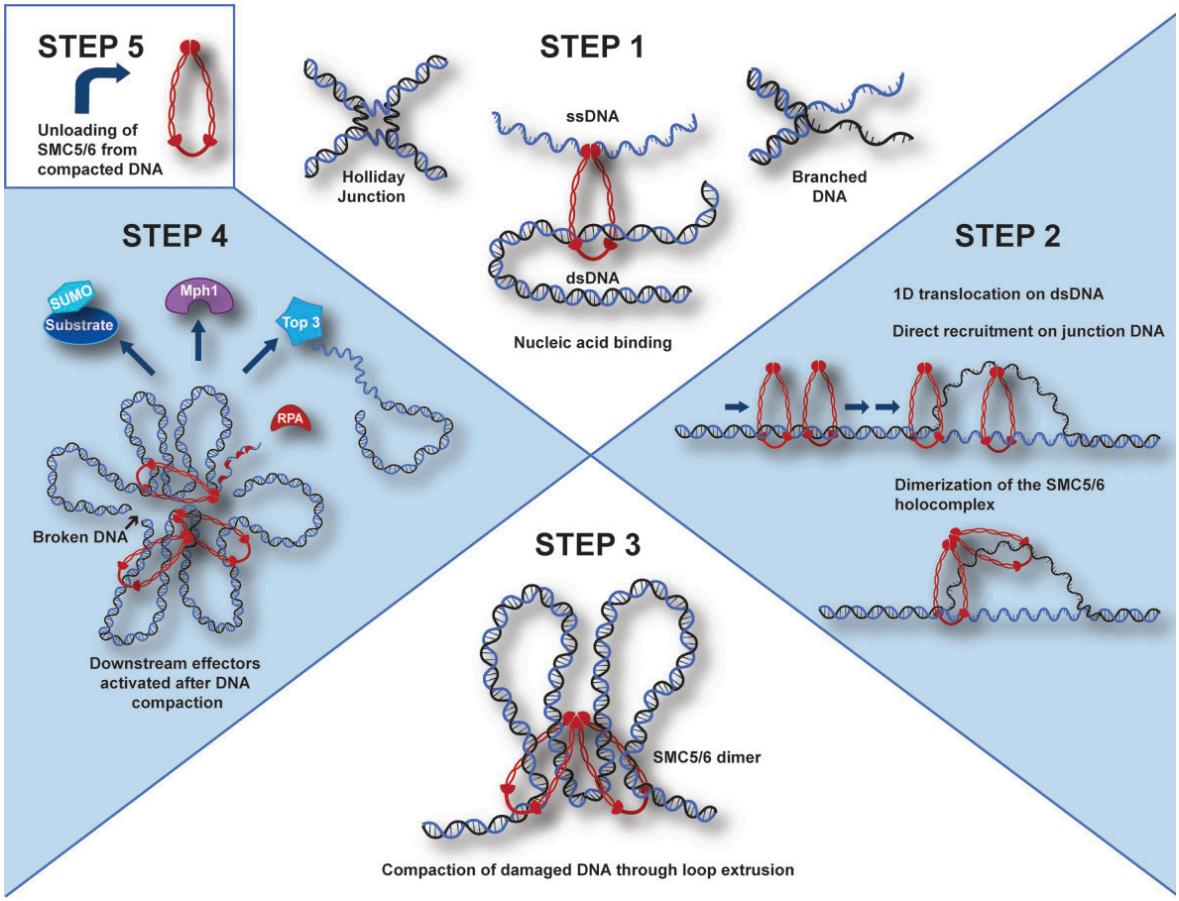


Figure 1. 5: A 5-step model depicting the catalytic cycle of the SMC5/6 complex on chromosomal DNA substrates. Step 1 shows the differential nucleic acid binding patterns of the SMC5/6 complex. Examples of relevant types of nucleic acid substrates (ssDNA, dsDNA, Holliday junctions, and branched DNA structures) are shown. The schematic illustration of the SMC5/6 complex binding to DNA represents a generic association and no specific mode of DNA binding should be inferred from the illustration (see text for more details). Step 2 illustrates the movement dynamics of the complex along DNA. After its initial binding to DNA, the SMC5/6 complex translocates on double-stranded DNA to scan for a relevant substrate. The SMC5/6 complex is enriched at junction DNA where it undergoes dimerization. Note that while dimerization of the complex has been established on undamaged DNA (Pradhan *et al*, 2023) and appears likely at sites of DNA damage, it nevertheless remains to be formally demonstrated in the context of DNA lesions. Step 3 shows the nature of DNA compaction carried out by the SMC5/6 complex. DNA compaction likely occurs through loop extrusion mediated by the dimeric complex. Step 4 proposes downstream effects of the SMC5/6 complex after DNA compaction is completed. Among these effects, we envision that the compacted DNA creates a protective environment as well as an effective platform for the recruitment of downstream effectors for DNA repair. Step 5 depicts the unloading of the SMC5/6 complex from the compacted chromatin. See main text for a more detailed explanation of this hypothetical model. Figure created using Biorender and Adobe Illustrator.

1.7 WHAT HAPPENS WHEN THE SMC5/6 COMPLEX FAILS?

Rare genetic disorders

Systematic identification of *de novo* and inherited mutations in subunits of the SMC5/6 complex have revealed its causative role in three rare genetic diseases. Notably, bi-allelic missense mutations in the human *NSMCE3* gene (encoding NSE3) have been associated with a fatal genetic disorder named lung disease immunodeficiency and chromosome breakage syndrome (LICS) (van der Crabben *et al*, 2016). NSE3 mutations disrupt the inter-subunit interactions within the complex ultimately leading to faulty HR, sensitivity to DNA replication stress, chromosomal rearrangements, and micronuclei formation (van der Crabben *et al*, 2016) (Figure 1.2).

Another report identified compound heterozygous frameshift mutations in the *NSMCE2* gene

leading to reduced NSE2 protein abundance in patients with primordial dwarfism, extreme insulin resistance and gonadal failure (Payne *et al*, 2014). Alterations in *NSMCE2* correlated with increased frequency of micronuclei, nucleoplasmic bridge formation, and DNA replication delay in patients (Payne *et al*, 2014) (Figure 1.2). Patients carrying mutations in the genes encoding SMC5 and its targeting factor SLF2 experience multiple mitotic abnormalities that give rise to a prognostic genome instability phenotype consisting of segmented, dicentric and rail-road chromosomes associated with mosaic variegated hyperploidy (MVH), a disorder known as the Atelís syndrome (Grange *et al*, 2022) (Figure 1.2).

The types of cellular phenotypes observed in the three syndromes listed above are fully consistent with the known functions of the SMC5/6 complex in DNA replication and repair processes. We anticipate that the hypomorphic mutations affecting other subunits of the complex are likely to occur at a low frequency in humans and would give rise to similar but still undiscovered inherited disorders in patients.

Cancer

Mutations in DNA repair enzymes have historically been associated with increased rates of cancer formation in humans (Jeggo *et al*, 2016; Perera *et al*, 2016; Helleday *et al*, 2014; Romero-Laorden & Castro, 2017). Consistent with this view, studies have reported the presence of gene alterations affecting SMC5/6 complex components in a number of human cancers, including: hepatocellular carcinoma (Nie *et al*, 2021; Yang *et al*, 2020), sarcoma (Zhou *et al*, 2020), breast cancer (Di Benedetto *et al*, 2022), colorectal cancer (Venegas *et al*, 2020) and brain metastasis (Saunus *et*

al, 2015) (Figure 1.2). Whether those alterations act as passenger mutations or driver of cancer formation remained unknown until recently. Experiments using a murine model of cancer have revealed that haplo-insufficiency of *NSMCE2* is associated with tumor formation and overall poor survival (Jacome *et al*, 2015), supporting the notion that loss of SMC5/6 complex function can promote oncogenesis.

A very recent phenogenomic analysis of cancer genomes suggests a direct link connecting SMC5/6 complex mutations to cancer development (Roy *et al*, 2023). This study evaluated the presence, frequency and nature of SMC5/6 gene alterations across a vast range of tumor types in a dataset comprising ~65,000 cancer genomes. This systematic analysis revealed that components of the human SMC5/6 complex are frequently altered in several types of cancers, including breast, prostate and ovarian cancers. Patients carrying genetic alterations in the subunits of the SMC5/6 complex experience strong phenotypic effects involving changes in genome ploidy and reduced overall survival compared to patients carrying a wild-type SMC5/6 complex (Figure 1.2).

In closing, we note that virus infections associated with degradation and/or inactivation of the SMC5/6 complex might also lead to tumorigenic conditions. This notion finds support in human cases of hepatitis B virus (HBV) infection where the viral regulatory protein X (HBx) induces the degradation of the SMC5/6 complex (Decorsière *et al*, 2016), leading to a reduced DNA repair capacity and accumulation of DNA damage in liver cells. Not surprisingly, this situation is associated with increased tumorigenesis in HBV-associated hepatocellular carcinoma (Funato *et*

al, 2022; Sekiba *et al*, 2022).

1.8 CONCLUSION

Taken together, research over the last few years has revealed that the SMC5/6 complex promotes the repair of eukaryotic genomes using a sequence of multimodal interactions that reconfigures chromosomal DNA in space. This mode of action is distinct from canonical pathways of chromatin modification that act at the level of histones, PTMs and chromatin remodeling enzymes. Recent structural and biophysical studies have provided a clear outline of the steps involved in the catalytic cycle of the SMC5/6 complex, but it will be important for future research to focus on revealing how the reorganization of chromatin through looping and compaction can specifically enhance DNA repair reactions as well as other genome transactions. Given the multifarious impact of the SMC5/6 complex on genome stability, future research on this topic is likely to provide remarkable mechanistic advances on DNA repair pathways and their health-related impacts in oncology, virology and aging.

1.9 PRIMARY OBJECTIVES OF THIS THESIS

In this thesis, I investigated novel functions and biological activities of the Smc5/6 complex essential for promoting genomic stability and preventing the development of chromosomal abnormalities. My primary objectives were as follows:

1. To explore if the SMC5/6 complex is altered in human cancers and assess the functional impact of cancer-associated mutations in the Smc5/6 complex subunits in yeast.
2. To investigate whether and how the Smc5/6 complex binds and resolves harmful nucleic

acid structures like RNA-DNA hybrids/R-loops during transcription and other biological pathways.

3. To identify novel inter-subunit interaction surfaces within the Smc5/6 complex which are required for its core biochemical activities.

1.10 MODEL ORGANISM

I have used *Saccharomyces cerevisiae* as a model organism for my studies. Yeast is a unicellular eukaryote, which can exist in both haploid (one copy of each gene) or diploid (two copies of each gene) conditions as well as undergo both asexual and sexual life cycles. Fundamental eukaryotic biology, such as cell cycle control and genome architecture are well-conserved throughout eukaryotic taxa including yeast. The budding yeast *Saccharomyces cerevisiae* has emerged as one of the most robust model systems for studying complex cellular processes conserved in eukaryotic evolution. Additionally, mutant screening and segregation analysis are simpler and easier to perform in yeast than in multicellular organisms. Yeast has naturally elevated rates of endogenous homologous recombination, and a host of extrachromosomal DNA elements can be transformed and stably maintained in yeast cells. Moreover, yeast is well suited for the development and use of temperature-sensitive mutations where the mutant alleles show a wild-type phenotype at a permissive temperature and a defective phenotype at its restrictive temperature. By exploiting the multiple advantages of yeast genetics, it is possible to study complex genetic interaction among various genes as well as study the effect of essential genes on genome stability and cell homeostasis (Duina *et al*, 2014; Hartwell, 1967; Herskowitz, 1988).

Chapter 2: Manuscript #1

Large-scale phenogenomic analysis of human cancers uncovers frequent alterations affecting SMC5/6 complex components in breast cancer

Published: NAR Cancer, Volume 5, Issue 3, September 2023, zcad047,
https://doi.org/10.1093/narcan/zcad047

Shamayita Roy[1], Arvin Zaker[2], Arvind Mer[2,3] & Damien D'Amours[1,3]

[1] Ottawa Institute of Systems Biology, Department of Cellular and Molecular Medicine,

[2] Department of Biochemistry, Microbiology & Immunology, University of Ottawa, Roger Guindon Hall, 451 Smyth Rd, Ottawa, ON, K1H 8M5, Canada

[3] Corresponding authors

Email: amer@uottawa.ca, Damien.Damours@uottawa.ca

2.1 ABSTRACT

Cancer cells often experience large-scale alterations in genome architecture because of DNA damage and replication stress. Whether mutations in core regulators of chromosome structure can also lead to cancer-promoting loss in genome stability is not fully understood. To address this question, we conducted a systematic analysis of mutations affecting a global regulator of chromosome biology –the SMC5/6 complex– in cancer genomics cohorts. Analysis of 64,959 cancer samples spanning 144 tissue types and 199 different cancer genome studies revealed that the SMC5/6 complex is frequently altered in breast cancer patients. Patient-derived mutations targeting this complex associate with strong phenotypic outcomes such as loss of ploidy control and reduced overall survival. Remarkably, the phenotypic impact of several patient mutations can be observed in a heterozygous context, hence providing an explanation for a prominent role of SMC5/6 mutations in breast cancer pathogenesis. Overall, our findings suggest that genes encoding global effectors of chromosome architecture can act as key contributors to cancer development in humans.

2.2 INTRODUCTION

Accurate replication, repair and segregation of genomic DNA are essential processes for normal cell growth and maintenance of genome stability. Failure to execute these processes accurately leads to genome instability and disease states in humans and other organisms (1). In particular, a compelling body of evidence indicates that numerical as well as structural alterations of chromosomes can act as key initiating events in the development of cancer and other severe illnesses (1–4). Thus, proteins that play significant roles in upholding genome stability –namely, DNA repair and cell cycle checkpoint proteins– are often the ones that are found to be mutated in the genomes of cancer patients (5, 6). It is perhaps surprising, however, that much less is known regarding the role of global effectors of chromosome structure/morphogenesis in cancer development, especially when considering the prevalence of chromosome-level changes in the genomes of cancer patients (*e.g.*, ~90% or more aneuploidy).

Structural Maintenance of Chromosome (SMC) protein complexes are responsible for the large-scale organization of the genetic material within cells (reviewed in refs 7–10). The SMC family includes the well-characterized cohesin (SMC1/3) and condensin (SMC2/4) complexes as well as the more recently identified SMC5/6 complex. In eukaryotes, cohesin maintains proximity and alignment of sister chromatids whereas condensin is associated with the compaction/morphogenesis of chromosomes during cell division (7, 10). Recent studies have shown that several cancer types harbor rare mutations in cohesin and condensin complexes supporting their role in suppression of tumorigenesis (11–13). Whether they contribute more broadly to oncogenesis remains to be established.

The SMC5/6 complex has important roles in diverse cellular processes affecting chromosome integrity throughout the cell cycle. Its main functions include the repair of DNA double strand break and stabilization of stalled replication forks (reviewed in refs 8, 9). Furthermore, the NSMCE2/Mms21 subunit of the SMC5/6 complex sumoylates multiple chromosome proteins via its E3 SUMO ligase activity (9, 14–16) and is involved in the alternate lengthening of telomeres (ALT) maintenance pathway (17, 18). We (and others) have found that the SMC5/6 core complex is a structure specific DNA binding enzyme (19–23) that promotes genome stability via its DNA compaction activity (24, 25). Due to its major role in upholding genomic integrity, misregulation of the SMC5/6 complex often result in abnormal cellular growth and severe chromosomal abnormalities (reviewed in refs 8, 9). This raises the intriguing possibility that alterations in the components of the SMC5/6 complex might be associated with cancer development and other tumorigenic processes in humans. So far, very few reports have explored whether misregulation of the SMC5/6 complex might be directly responsible for cancer development in humans (17, 26). The components of the SMC5/6 complex have been shown to be upregulated or altered in some cancer types (for example, hepatocellular carcinoma (27, 28), sarcoma (29), breast cancer (30), and brain metastasis (31)). However, it is currently unknown if these changes are passenger mutations or play a greater role in triggering oncogenesis in patients. This highlights the need for systematic studies to better define the significance of SMC5/6 complex alterations in cancer etiology and patient survival.

To fill this knowledge gap, we performed an integrative analysis of cancer genomic data (32) to identify the frequency and nature of cancer-associated alterations in the components of the SMC5/6 complex. Our phenogenomics analyses revealed that genes encoding the subunits

of the SMC5/6 complex are frequently altered in breast as well as other types of human cancers. Importantly, these alterations are associated with severe ploidy aberrations and lower overall survival rates in patients. Extensive validation of the phenotypic effects associated with genetic variants of unknown significance (VUS) revealed a range of mutational severity, from mild to severe growth defects for point mutations affecting conserved functional domains, to lethality with alleles that truncate SMC5/6 components. Our results unravel a previously underappreciated pattern of frequent alterations in the components of the SMC5/6 complex in cancer genomes.

2.3 MATERIALS AND METHODS

Cancer cohort analysis

Cancer genomics and patient survival data was obtained from the cBioPortal for Cancer Genomics (https://cbioportal.org) web resources on August 2022 (32). Obtained data contained 64959 samples from 199 studies and 145 tissue types (577 tissue subtypes). During the quality control steps, duplicated samples, pediatric cancer cohorts and cohorts with samples size <20 were removed. Next, we selected patient samples for which mutation, copy-number alteration (CNA), and structural variation (SV) information was available. The final dataset contains 29316 samples from 33 tissue types and 100 studies. Data processing, analysis and visualization was performed on R statistical software (version 4.2.2) using packages ggplot2 (version 3.4.0), ComplexHeatmap (version 2.14.0) and circlize (version 0.4.15). In the analysis, the SMC5/6 complex was considered altered if any of the genes encoding its constituents (*NSMCE1*, *NSMCE2*, *NSMCE3*, *NSMCE4A*, *SMC5*, *SMC6* and *EID3*) had alteration. Survival analysis was performed using R package survival

(version 3.4-0) and survminer (version 0.4.9). Kaplan-Meier plots were used for visualization of survival differences and log-rank-test was used to estimate statistical significance.

Variation allele frequency analysis of patients with SMC5/6 complex mutations was performed on the TCGA dataset. Mutation allele frequency files were downloaded using the TCGAbiolinks (version 2.25.3) package (33) for the samples containing point-mutation in SMC5/6 complex components. We then calculated the variation allele frequency (VAF) using the formula:

$$VAF = \frac{f_{altered}}{f_{altered} + f_{reference}}$$

Where $f_{altered}$ is the frequency of the DNA sequence reads with the genomic alteration, and $f_{reference}$ is the frequency of the reference DNA sequencing reads without genomic alterations. A variant is generally considered to be homozygous if its VAF is above 50%, meaning that more than half of the sequencing reads support the variant allele. We then visualized the VAF values of all patients with recorded SMC5/6 complex alteration using ggplot2 (version 3.4.0).

Differential gene expression and pathway analysis

Differential gene expression analysis of microarray data was performed using limma package (version 3.52.4) (34). For RNAseq based gene expression data, edgeR package (version 3.40.0) (35) was used, and data were fitted using the GLM method (35). Gene expression data for METABRIC (36) study was obtained using the MetaGxBreast package (version 1.18.0) (37). STAR aligner (38) based The Cancer Genome Atlas (TCGA) (39) RNA-Seq data was downloaded using the TCGAbiolinks package (version 2.25.3) (33).

Gene ranking based on the log fold change values were utilized for gene set enrichment analysis. We obtained the canonical pathways (version 7.5.1) from the Molecular Signatures

Database (MSigDB). Pathways with number of genes <25 and >150 were removed from the analysis. R package fgsea (version 1.25.0) was used for gene set enrichment analysis and statistical significance (p-value) for each pathway was estimated using adaptive multi-level split Monte-Carlo scheme (40). False discovery rate (FDR) was used for p-value adjustment of multiple hypothesis testing. Pathways were categorized into groups such as DNA replication and repair, cancer-specific, cell cycle-related, metabolic, signaling, extracellular matrix, immune pathway based on semantic analysis (41). Enriched pathway network was constructed where node represents pathways and edge represents number of shared genes between pathways. Nodes were connected only when more than 20 genes were shared between nodes. Visualization of network was done using igraph package (version 1.3.5).

Comparative analysis of SMC5/6 alterations with known effectors of genomic instability

Co-mutational pattern of SMC5/6 complex was analyzed against several genomic stability related genes. Mutation, copy-number alteration (CNA), and structural variation (SV) information for these genes were downloaded from the cBioPortal web server. We then compared the co-mutational patterns of the SMC5/6 complex against the genomic instability genes using the co-mutation frequency table. For the analysis we computed false discovery rate (FDR) (42) of co-mutation pattern. FDR was defined as:

$$FDR = \frac{FP}{TP + FP}$$

Where FP is false positive (number of samples with alteration in genomic instability gene and wild-type SMC5/6 gene), and TP is true positive (number of samples with alteration in both

genomic instability and SMC5/6 genes). This analysis was performed for each SMC5/6 complex gene and genomic stability gene pair. Pie chart was used for co-mutation frequency visualization.

We also compared the transcriptional signature associated with SMC5/6 complex alteration against genomic instability related transcriptional signatures. For this analysis, we first performed differential gene-expression analysis between samples with high and low genomic instability. If at least three of 12 essential genomic stability genes were mutated, cancer samples were labeled as exhibiting high genomic instability. Log fold change (LFC) values associated with genomic stability were compared to LFC values associated with SMC5/6 complex changes for all genes.

Mutation profile analysis

Lollipop plots of different mutation profiles for respective genes in all cancers were generated by the mutation mapper in cBioPortal. Evolutionary conservation of the mutations in SMC5/6 complex subunits were compared in 5 species, *Saccharomyces cerevisiae* (budding yeast), *Saccharomyces pombe* (fission yeast), *Xenopus laevis* (frog), *Drosophila melanogaster* (fruit fly), *Homo sapiens* (human). The amino-acid sequence of eukaryotic homologs of SMC5/6 complex subunits were acquired from Uniprot (https://www.uniprot.org/). The amino acid sequences of the genes were aligned in the Clustal Omega sequence alignment tool (https://www.ebi.ac.uk/Tools/msa/clustalo/). Additionally, the conserved amino-acid residues were shaded with Boxshade online tool (https://embnet.vitalit.ch/software/BOX_form.html).

Yeast strains and cell viability assay

All yeast strains used in this study are derivatives of strain K699/K700. The genotype of the yeast strains used in the study are listed in supplementary information (Table 1.S1). Yeast growth conditions, media composition and procedures for genetic analysis can be found elsewhere (43). For experiments performed under conditions of DNA damage or replication stress, yeast cultures were grown on solid medium containing methyl methanesulfonate (MMS), 4-nitroquinoline 1-oxide (4-NQO) and hydroxyurea (HU) at 23°C, 30°C and 37°C. Temperature and DNA damage sensitivity were monitored for all mutants created in this study because several DNA repair/homologous recombination mutants show sensitivity to both types of stress (44, 45). In summary, 5-fold dilution series of wild-type and mutant yeast cultures (first spot on the left side of the plate corresponds to a culture at OD_{600} of 0.2) were spotted on solid YPD (yeast extract, peptone, 2% glucose) and grown in temperature-controlled incubators for 48-72 hours before scanning the plates in a scanner (46). All the experiments were repeated a minimum of 3 independent times, and we show representative results in figures.

Plasmid and mutant construction

Yeast strains expressing cancer-specific gene truncations were created by integration of a *T_{ADH1}-kanMX6 or T_{ADH1}::URA3MX6* cassette at the desired point of truncation in the endogenous loci of genes encoding the SMC5/6 complex subunits. Heterozygous diploid strains carrying these truncation alleles were then sporulated on minimal media and the viability of the dissected haploid spores was determined after 4 days of growth on solid medium. Missense mutations were initially introduced in plasmids (*pFA6a-ORF::T_{ADH1}::kanMX6* or *pFA6a-ORF::T_{ADH1}::URA3MX6*) carrying *SMC5*, *SMC6*, *NSE1*, *MMS21*, *NSE3*, or *NSE4* genes using

QuikChange Multi Mutagenesis kit (Agilent). Mutant alleles of the respective genes were then transformed into wildtype diploid yeast strain at the relevant endogenous loci by transformation of a PCR product from the region of interest in the respective mutagenized plasmids. Heterozygous diploid strains carrying these missense mutation alleles were then sporulated and dissected to determine the phenotype of the mutant haploid strains. All the mutations in the haploid mutants were confirmed by sequencing the respective genetic loci.

Immunoblot analysis

Cell lysates were prepared from exponential cultures of yeast grown at 23 °C or 37 °C using the TCA glass-bead method (47). Lysates were subsequently resolved by SDS-PAGE and processed for immunoblot analysis using anti-Mcm4 antibody (clone C-12; 1:500 dilution; Santa Cruz Biotechnology; sc-166036), anti-Myc antibody (clone 9E10; 1:2500 dilution; Cedarlane; GTX20032), anti-Pgk1 antibody (clone 22C5D8; 1:2500 dilution; Abcam) and an anti-mouse IgG antibody (1:5000 dilution; Cytiva). Band intensity on immunoblots was measured using Adobe Photoshop (version 2022). Data are presented as means ± SEM. All statistical analyzes were performed using GraphPad Prism 7 (GraphPad Software Inc) and statistical significance threshold was set at p-value = 0.05. Where indicated in figure legends, we performed multiple t-tests comparing the mutants to the wild-type for either temperature, p-values were corrected by Bonferroni multiple correction method.

2.4 RESULTS

Subunits of SMC5/6 complex are frequently altered across different cancer types

The components of the human SMC5/6 complex assemble into a ring-like complex similar in configuration to that of cohesin and condensin, the defining members of this family of proteins (8, 9) (Figure 2.1A). The ring structure of the complex is formed by the dimerization of two SMC proteins, SMC5 and SMC6, and their association with a group of accessory subunits named NSMCE1, NSMCE3 and NSMCE4A/EID3. Additionally, the human SMC5/6 complex comprises an E3 SUMO ligase, NSMCE2 (also known as Mms21/Nse2) (14, 15), which interacts with the coiled-coil domain of SMC5 (Figure 2.1A) (48, 49).

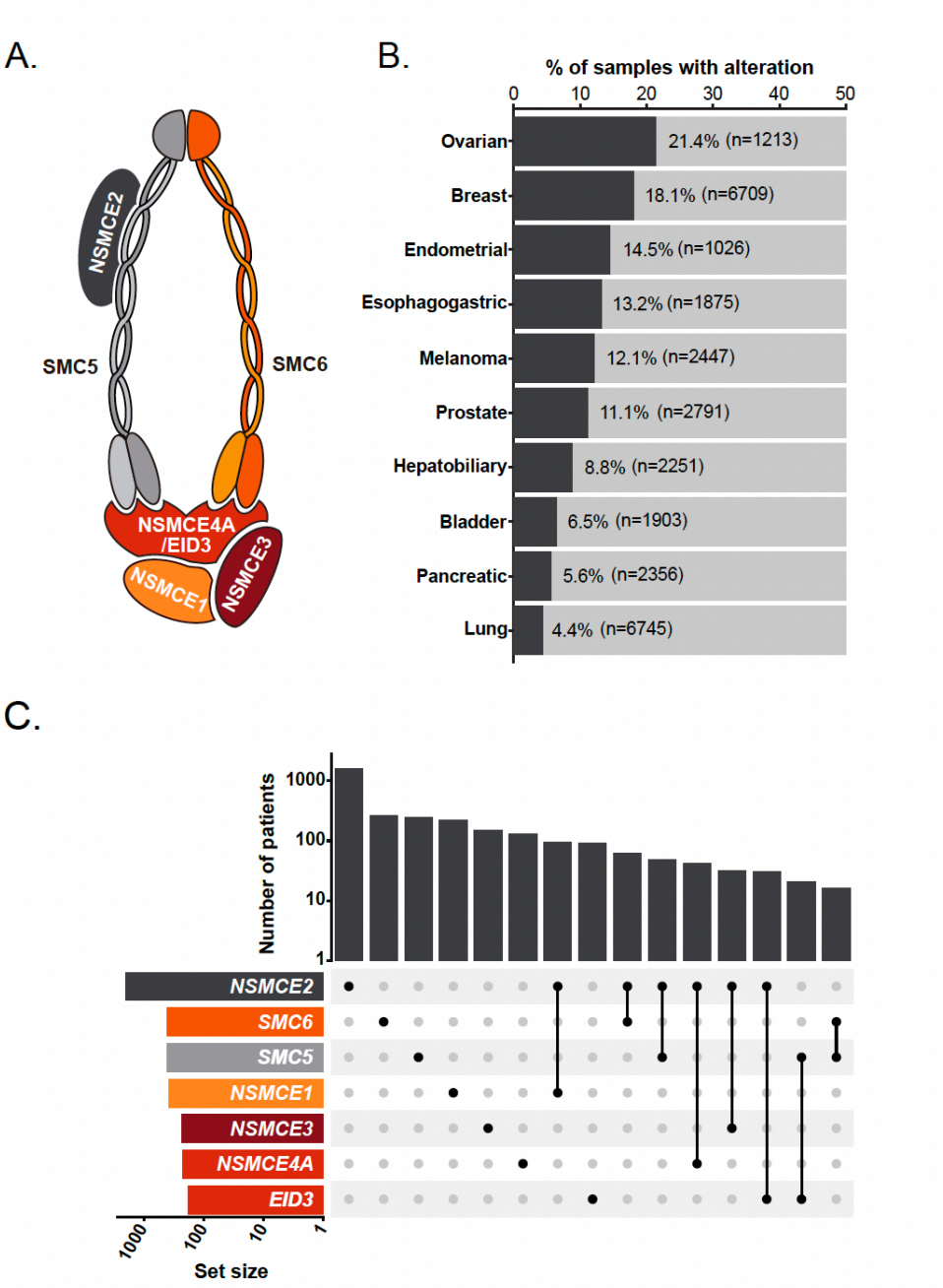


Figure 2. 1: Genetic alterations affecting the subunits of the SMC5/6 complex in human cancers. **(A)** Schematic representation of the SMC5/6 complex. Note that the NSMCE4A subunit is replaced by EID3 in male gonads (50). **(B)** Bar chart showing the frequency of SMC5/6 complex alteration in cancer. For visualization, 10 tissue types with the highest frequency of SMC5/6 complex alteration were chosen. **(C)** UpSet plot of the distribution of genomic alteration of each subunit of the SMC5/6 complex in the 10 major cancer tissue types reported herein. Number of patients (y-axis) are shown using logarithmic scale.

We analyzed 64,959 cancer samples spanning across 144 tumor types and 199 different cancer genome studies for evidence of genetic alterations in the 7 genes encoding the known components of the SMC5/6 complex (Figure 2.1A). Genomic alterations including structural variation (SV), mutations, and copy-number alterations (CNA), at the SMC5/6 complex were investigated. After data preprocessing and quality control, we derived 29,316 high-quality cancer sample profiles across 34 tumor types including lung cancer (N=6,745), breast cancer (N=6,709), and prostate cancer (N=2,791) (Figure 2.1B and Supplementary Figure 2.S1). Analysis of the curated dataset revealed that genomic alteration of the SMC5/6 complex is present in 10.9% of all cancer samples.

Stratifying by tissue of origin revealed that genetic alterations affecting SMC5/6 complex components are most prevalent in ovarian cancer (21.4%) followed by breast (18.1%) and endometrial cancer (14.5%). Analysis of cooccurrence patterns revealed that the majority of patients have alteration in only one gene encoding members of the SMC5/6 complex and *NSMCE2/NSMCE1* have high frequency of coexisting mutation (Figure 2.1C and Figure 2.2). The general trend of SMC5/6 subunit alterations we observed in individual cancers (Figure 2.2) is similar to the trend we observed in aggregate (Figure 2.1). Namely, amplification of *NSMCE2* is the most common genomic alteration, except in endometrial cancer where missense mutations are prevalent (Supplementary Figure 2.S1). Genomic alteration of *NSMCE2* is dominant in breast, prostate, and ovarian cancer while alteration in *SMC5* and *SMC6* genes are highly prevalent in lung, melanoma, bladder, and endometrial cancer (Figure 2.2). To determine the zygosity of SMC5/6 complex mutations, we conducted a variant allele frequency (VAF) analysis of cancer alterations affecting the subunits of the complex. Only 0.4% of samples had biallelic mutations

(VAF > 0.5) in any of the SMC5/6 complex subunits, showing that the vast majority of SMC5/6 complex mutations are heterozygotes (Figure 2.3A; one tail T-test p-value < 0.0001).

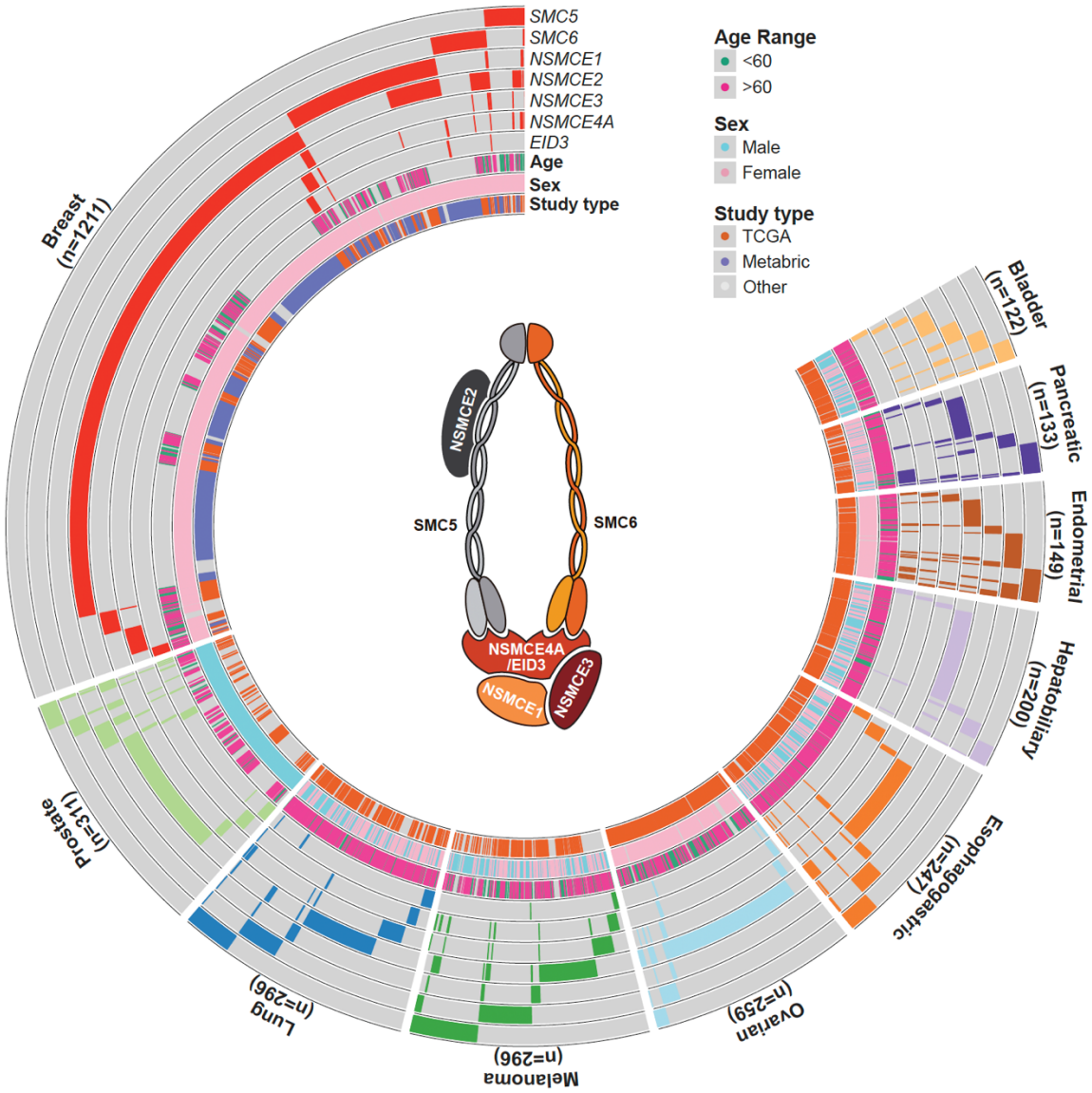


Figure 2. 2: Circular oncoprint of patients carrying genomic alteration in the SMC5/6 complex. Each patient is represented by a cross-section of the circle. Recorded genomic alteration in each of the subunits is shown as a color relating to the cancer type. Age is dichotomized at 60 years. The number of patients with genomic alteration in SMC5/6 complex belonging to each cancer type is shown under the name.

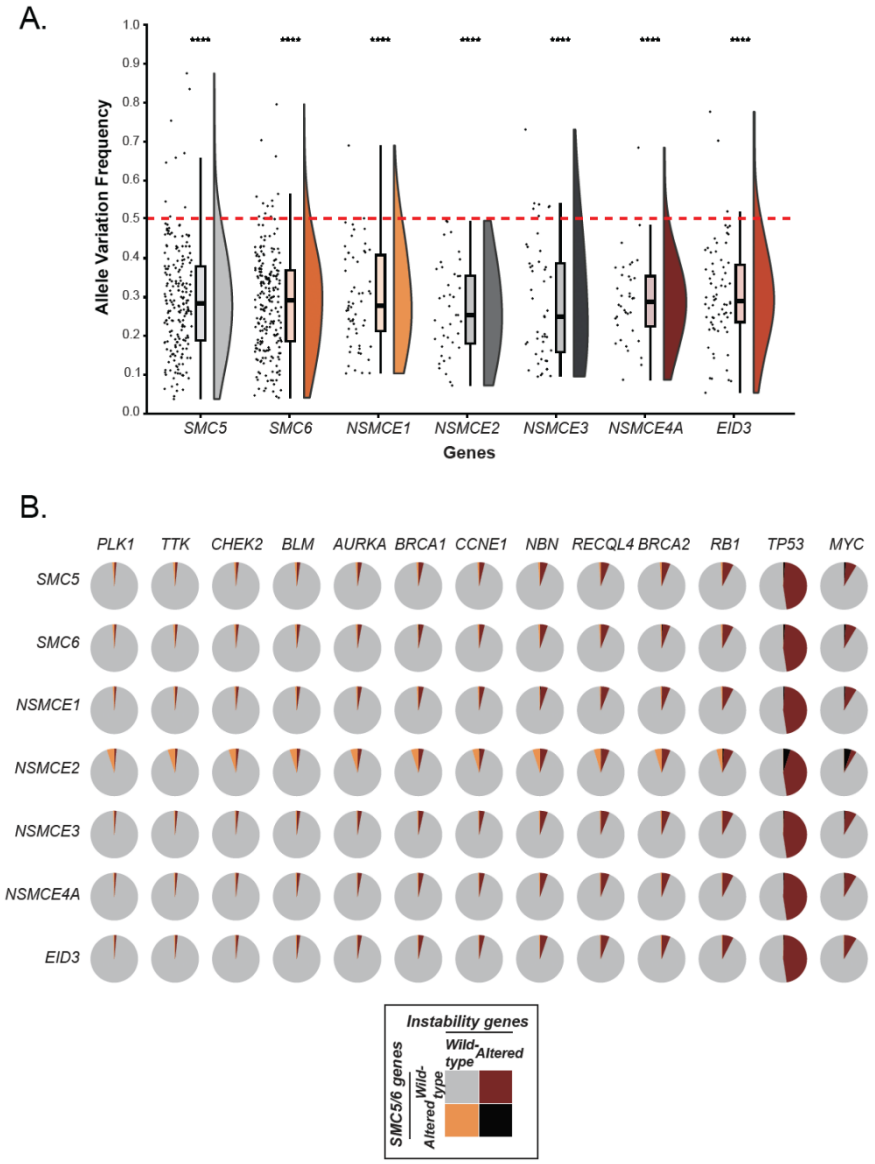


Figure 2. 3: (A) Zygosity of SMC5/6 complex alterations in human cancers. Different SMC5/6 complex genes are shown on the X-axis of the graph, whereas the Y-axis shows variant allele frequency (VAF) in cancer samples. Most mutations have VAF < 0.5 indicating that they are heterozygotes. Statistical significance was computed using one tail t-test (H1: average VAF is < 0.5), and p-value < 0.0001, represented by **** at the top of each gene. **(B)** Co-mutation patterns between SMC5/6 complex and genes associated with genomic instability. Each row represents a component gene of the SMC5/6 complex, while columns represent gene related genomic instability. The mutational patterns are displayed as a pie chart, with the percentage of samples that have a wild-type status in both genes represented in the color gray, and the frequency of samples that have a mutation in both genes shown in the color black. The color orange indicates a mutation that has only occurred in the SMC5/6 complex, whereas the color dark red indicates a mutation that has only occurred in the genomic instability gene.

Next, we evaluated whether mutations in the SMC5/6 complex are directly linked to mutations in genomic stability-related genes (Figure 2.3B). We analyzed co-mutations patterns between genes encoding the SMC5/6 complex and key genes associated with genome stability and computed false discovery rate (FDR) for each combination. We did not observe statistically significant co-mutation patterns (all FDR > 0.05). We discovered that the *TP53-SMC5/6* complex co-mutation rate was the highest. This was expected (and not statistically significant) given the high prevalence of *TP53* mutation in cancer. These results strongly suggest that genetic alterations affecting the SMC5/6 complex are not directly driven by mutations in genomic stability-related genes.

Genomic alteration in SMC5/6 complex is associated with aggressive disease

To understand the effect of SMC5/6 complex alterations on aggressive cancer, we performed an in-depth analysis of its association with ploidy score and survival outcome (Figure 2.4). We focused on breast cancer, as it has the highest number of samples with the SMC5/6 complex alterations (Figure 2.2). Analysis of breast cancer genomes revealed that individuals harboring the SMC5/6 complex alteration have significantly higher aneuploidy scores (Figure 2.4A, Wilcoxon signed-rank test p-value = 3×10^{-4}) and ploidy scores (Figure 2.4B, Wilcoxon signed-rank test p-value = 1×10^{-6}). This association with high aneuploidy and ploidy scores was also observed in cancer genomes from individuals carrying *NSMCE2* alterations. Analysis of male-specific prostate cancer data showed a similar trend (Supplementary Figure 2.S2), suggesting that high aneuploidy and ploidy scores are likely due to SMC5/6 complex alterations. Survival analysis using Kaplan-Meier (KM) plots shows that patients with altered SMC5/6 complex had significantly

lower overall survival rates compared to patients with normal SMC5/6 complex components (Figure 2.4C, Log-rank p-value = 6.0×10^{-4}). We also observed that the *NSMCE2* alteration is associated with significantly worse patient survival relative to individuals carrying wild-type *NSMCE2* (Figure 2.4D, Log-rank p value = 1.0×10^{-5}).

We observed a consistent pattern while stratifying the genomic alterations into categories. The amplification of SMC5/6 complex genes were associated with higher aneuploidy and ploidy score and poor patient survival (Supplementary Figure 2.S3). The same holds true for *NSMCE2* gene amplification. Point mutation to ploidy score analysis was not possible due to a lack of data. However, survival analysis unequivocally demonstrates that mutations in the SMC5/6 complex result in worse prognosis, even when compared to amplification (Supplementary Figure 2.S3B). The difference in overall patient survival in cohorts of patients carrying wild-type and altered SMC5/6 complex components was also observed in individuals with prostate and ovarian cancer (Supplementary Figures 2.S2 and 2.S4, respectively). Overall, our analyses show that the alterations in the genes encoding the SMC5/6 complex are linked to high cellular ploidy and are associated with poor prognosis with *NSMCE2* being the most frequent target of alteration.

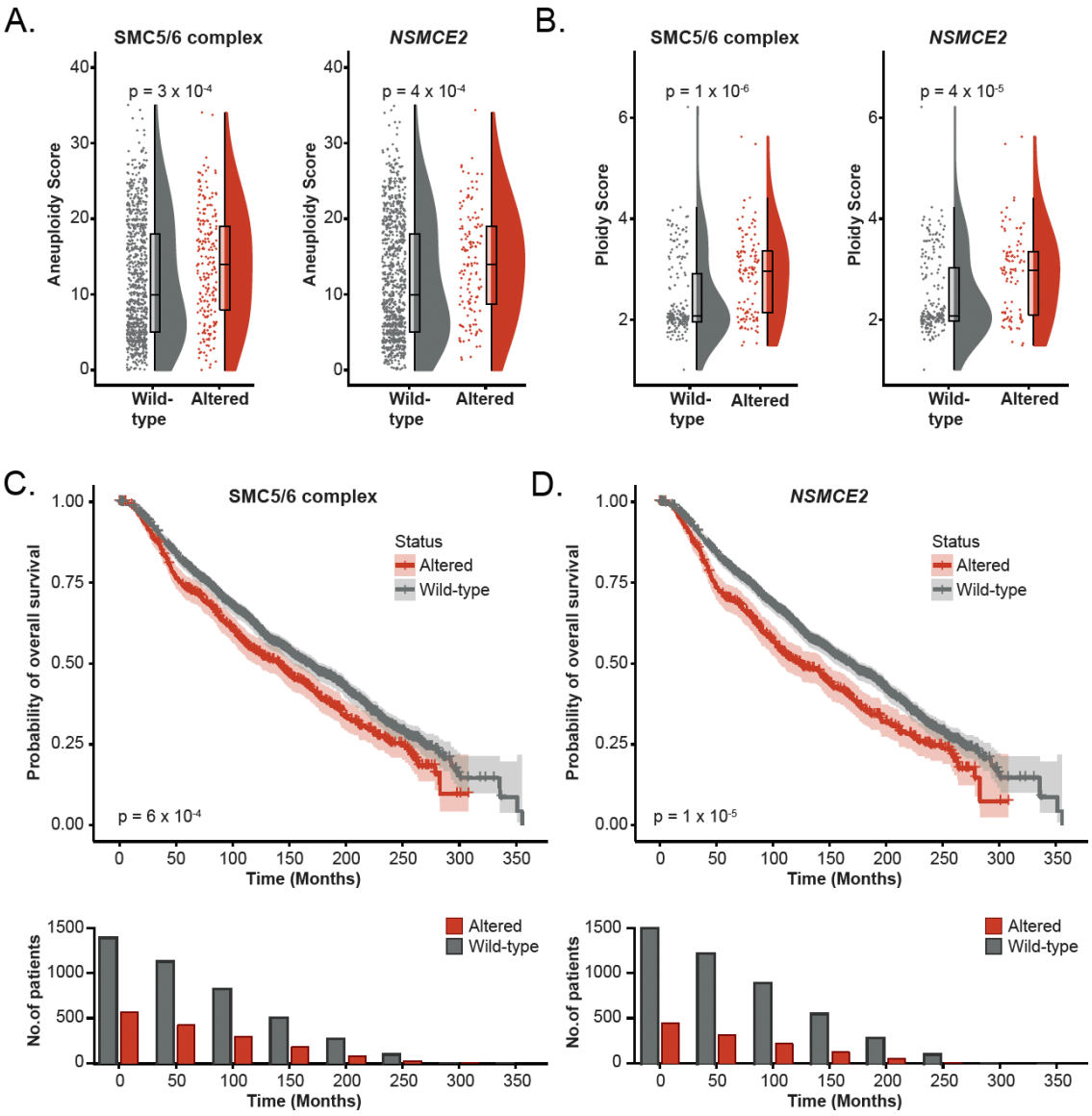


Figure 2. 4: SMC5/6 complex alterations are associated with genome stability defects and reduced survival in breast cancer patients. Distribution of (A) aneuploidy and (B) ploidy score in breast cancer patients. Data are stratified by SMC5/6 complex and NSMCE2 alteration status. Samples carrying SMC5/6 complex alterations have higher aneuploidy and ploidy score (two-sided Wilcoxon rank-sum test $p < 0.05$). Kaplan-Meier (KM) plots of overall survival in breast cancer, stratified by (C) SMC5/6 complex and (D) NSMCE2 gene alteration status. Altered samples are shown in red and wild type in gray. The bar charts at the bottom of the KM plots represent the number of patients at risk. Alteration in SMC5/6 complex leads to poor overall survival (Log-rank test p-value < 0.05).

SMC5/6 complex alteration are linked to DNA damage and replication stress

Differential gene expression analysis was performed to discover genes that are linked with SMC5/6 complex alteration in breast cancer (Figure 2.5A, and Supplementary Figure 2.S5). We found that several key genes involved in DNA damage repair, cell cycle and DNA replication were differentially expressed in the SMC5/6 complex alteration group (Figure 2.5B). Gene *RAD54B*, a member of DEAD-like helicase superfamily was upregulated in SMC5/6 complex alteration group (Figure 2.5B). It is known to play an active role in DNA damage repair and homologous recombination during cell division (51–53). *CCNE2* (Figure 2.5B), a cyclin family gene and vital component of G1/S transition during the cell cycle, showed high expression in SMC5/6 complex altered samples (54, 55). Similarly, *RECQL4*, and *MCM4* (Figure 2.5B) were upregulated in SMC5/6 complex alteration group. These genes are associated with DNA repair (56) and replication (57) respectively. Genes such as *KIF13B*, *PPP2R2A*, *PARP3*, and *TP53BP1* were all significantly downregulated in SMC5/6 complex alteration group (Figure 2.5B). Interestingly, these genes have been implicated in the negative control of cell growth (58), maintenance of genomic stability (32) and tumor suppressor activity (59). Therefore, downregulation of these genes may lead to aggressive tumor phenotype.

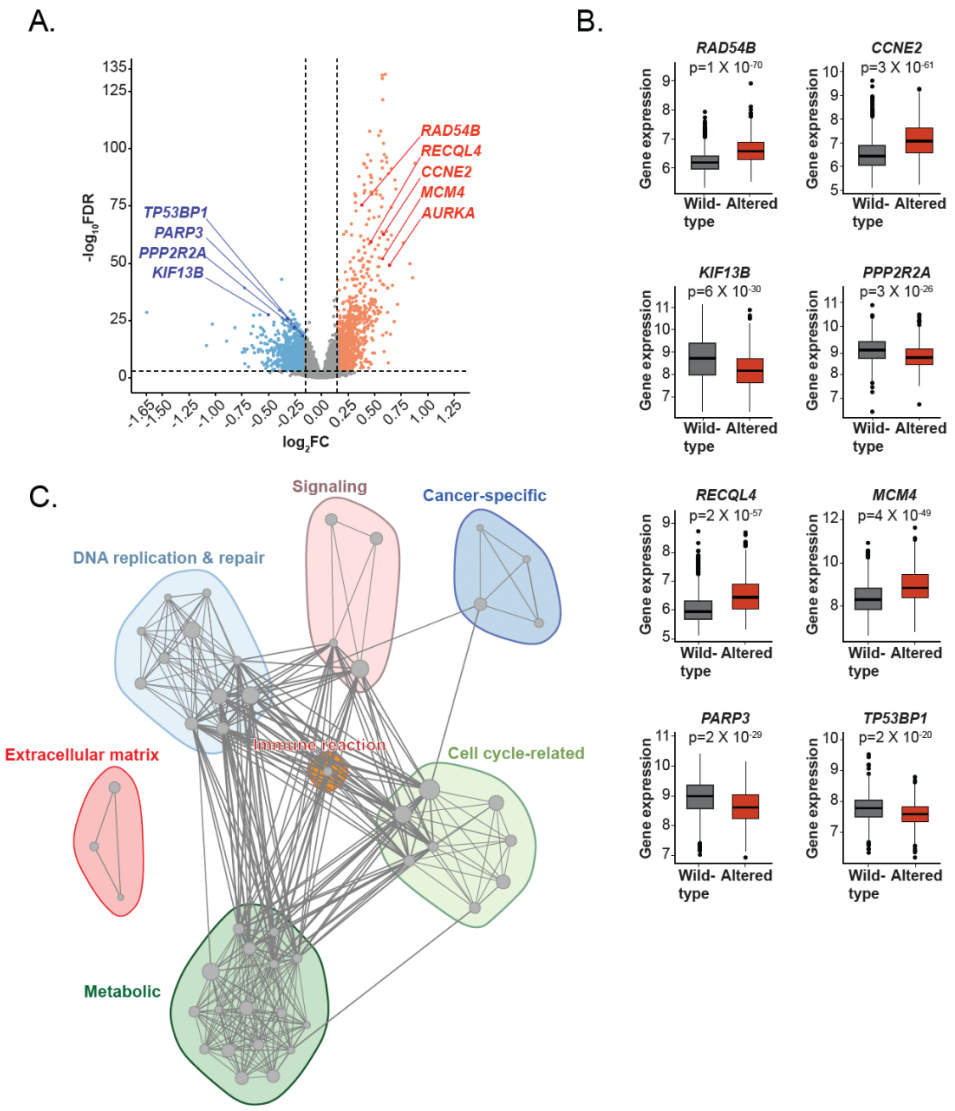


Figure 2. 5: Genomic alteration in the SMC5/6 complex leads to changes in gene expression and pathway activity. (A) Volcano plot of the differential gene expression analysis of breast cancer data. In the volcano plot x-axis represents log fold change (LFC) and y-axis represents false discovery rate (FDR) for genes. Significantly upregulated genes are highlighted with orange and significantly downregulated genes are highlighted with blue. (B) Box plot showing the expression of selected genes in SMC5/6 complex altered (red) and wild-type (gray) breast cancer samples. Differential gene expression p-value is shown on the top. (C) Network plot of enriched pathways in breast cancers carrying altered SMC5/6 complex components. Each pathway is represented by a node. The size of the node is proportional to pathway size. Edges between two nodes indicate the number of genes shared between pathways. Only links with more than 20 shared genes are shown. Pathways are categorized into groups using semantic analysis and network layout is done using Fruchterman-Reingold algorithm. Detailed pathway information is available in Supplementary Figure 2.S6.

We also compared the transcriptional signature of SMC5/6 complex modification to the transcriptional signature associated with genomic stability (Supplementary Figure 2.S5A). We found that only 84 genes were differentially expressed in both the conditions (*i.e.,* SMC5/6 complex alteration and mutation in key genomic stability genes). This constitutes only 3.72% of all genes that were differentially expressed in SMC5/6 complex. Consequently, 96.28% of the differentially expressed genes in SMC5/6 complex are unique to it. Furthermore, key genes such as *RAD54B*, *RECQL4*, and *MCM4* were uniquely linked to SMC5/6 complex alteration. Overall, these results established that the transcriptional changes associated with SMC5/6 complex alteration are independent of the genomic architecture abbreviations.

To explore the pathways associated with SMC5/6 complex alteration, we utilized gene-set enrichment analysis (GSEA) approach. Results revealed that the DNA replication and repair, cell cycle, metabolism, signaling and cancer related pathways are strongly associated with SMC5/6 complex alteration (Figure 2.5C and Supplementary Figure 2.S6). A similar trend was observed in TCGA prostate cancer data (Supplementary Figure 2.S5C). Taken together, these gene and pathway analyses results highlight a strong link between SMC5/6 complex alteration and DNA damage-related processes.

A large number of cancer mutations affect key functional domains of the SMC5/6 complex

In depth analysis of patient cancer genomes included in the study revealed that they harbor a large number of point mutations in the subunits of the SMC5/6 complex (Figure 2.6A). An analysis of the spatial distribution of mutations in the subunits of the SMC5/6 complex identified a total of 343 and 326 mutation sites in SMC5 and SMC6 proteins, respectively, with no major hotspot

region in either of the proteins. R972* nonsense mutation was the most frequently observed alteration in the SMC5 subunit in multiple cancers. Also, the I410Yfs frameshift deletion in SMC6 was found to be ubiquitously present in different cancer types (Figure 2.6A). Although the non-SMC subunits of SMC5/6 complex were also mutated at several sites in different cancer types, no single mutation were recurring at a higher frequency compared to the SMC core subunits. Importantly, several of the mutations identified in the SMC5/6 core complex and the non-SMC subunits were present on amino-acid residues which are evolutionarily conserved across several eukaryotic species (Figure 2.6B). Furthermore, in many cancers missense mutations affect domains and/or structural modules of known functional relevance for the SMC5/6 complex (e.g., H187Y in the SP-RING domain of NSMCE2 and R229G/Q in the WH-B domain of NSMCE3) (Figure 2.6B). Although most, if not all point mutations reported here should be viewed as variants of unknown significance (VUS), it stands to reason that VUS affecting evolutionarily conserved regions of the SMC5/6 complex are likely to disrupt the activity of the complex. Since we do not know which exact biochemical function of the SMC5/6 complex is affected by the cancer mutations described above, it is important to investigate the functional impact of each mutation on SMC5/6 complex activity.

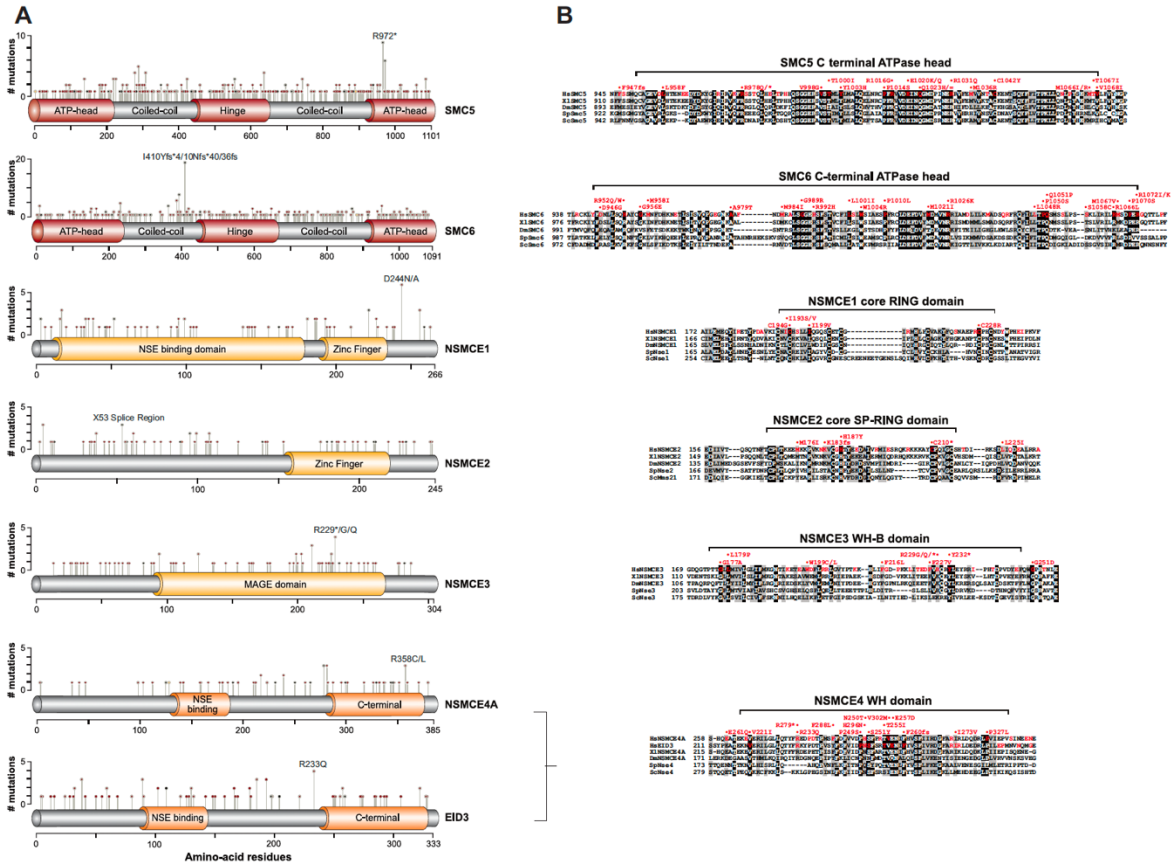


Figure 2. 6: Cancer mutations in the genes encoding components of the SMC5/6 complex. **(A)** The relative positions of cancer mutations are marked on schematics depicting the domain organization of the subunits of the SMC5/6 complex. Specific protein names are marked on the right-hand side of the schematics. The length of the line connecting the annotated mutation to the protein is indicative of the number of samples that have the mutation (frequency). The frequency of mutations is reported on the top of protein schematics. For each subunit, we also report the specific nature of the mutations most frequently identified in cancers. Specifically: Black circles mark the positions of for truncation mutations, red circles are for missense mutations and yellow circles are for frameshift mutations. **(B)** Position and evolutionary conservation of cancer mutations identified in functional domain of the SMC5/6 complex. Fully conserved residues are highlighted in black whereas positions conserved at ≥80% are highlighted in red. Mutations affecting positions that are evolutionarily conserved are marked in bold/red typeface, while mutations that affect non-conserved residues are marked in red typeface. Specific details pertaining to cancer mutations are marked above protein alignments. Mutational data are from cBioPortal. SMC5 C-terminal ATPase head region, SMC6 C-terminal ATPase head regions, NSMCE1 core RING domain, NSMCE2 core SP-RING domain, NSMCE3 winged helix B domain, and NSMCE4 winged helix domain.

Modeling cancer VUS reveals the impact of cancer mutations on SMC5/6 complex activity

We next sought to assess the functional relevance of cancer VUS on the activity of the SMC5/6 complex. We took advantage of the high evolutionary conservation of the SMC5/6 complex in eukaryotes to model the impact of point mutations in the budding yeast *Saccharomyces cerevisiae*. To achieve this, we selected a group of cancer VUS affecting conserved amino acid residues at several positions within the subunits of the SMC5/6 complex, focusing on the known functional domains of each subunit (Figure 2.6). We then constructed a series of heterozygous diploid yeast strains carrying cancer VUS at their endogenous loci and uncovered the phenotype associated with the mutations of interest after sporulation and dissection of haploid spores (Figure 2.7 and Table 2.S2).

The majority of the haploid yeast strains carrying truncating VUS in the components of the SMC5/6 complex were non-viable after sporulation, as evidenced by the 2:2 lethality phenotype co-segregating with each truncation allele on dissection plates (Figure 2.7A). The only viable truncation mutant obtained in this analysis, *mms21-C221**, showed severe proliferation defects at 23°C and 30°C (relative to a wild-type control strain) and was completely defective for growth at 37°C (Figure 2.7B). Even at its permissive temperature of 23°C, the yeast strain carrying *mms21-C221** was unable to grow effectively on media containing DNA damaging agents (methyl methanesulfonate [MMS] and 4-nitroquinoline 1-oxide [4NQO]) or a DNA replication inhibitor (hydroxyurea [HU]; Figure 2.7B). The inability of the *mms21-C221** mutant strain to withstand DNA lesions as well as its temperature-sensitive growth defect are highly similar to that of a known DNA repair defective mutant of the complex, the *smc5-6* mutant (Figure 2.7B) (60).

Interestingly, all haploid yeast strains carrying missense mutations were viable after sporulation (Figure 2.7C) but showed variable kinetics of proliferation under standard and DNA damage conditions. While all the mutants grew normally on solid medium (YPD) at 23°C, *mms21-H202Y* and *nse1-C321R* mutants appeared to be temperature sensitive at 37°C and showed poor proliferation on media containing MMS or HU (Figure 2.7C). Likewise, the s*mc6-G1021R* mutant was defective for growth at all temperatures tested when exposed to HU or MMS. Although yeast strains carrying the *smc5-E1015K*, *nse3-K236G* and *nse4-P315S* alleles showed normal growth phenotypes at the permissive temperature of 23°C, their proliferation was severely hindered when challenged with HU at the temperature of 37°C (Figure 2.7C). In summary, most of the mutant alleles we tested manifested defects in the functional activity of the SMC5/6 complex that resulted in mild to severely impaired growth phenotypes *in vivo.* A summarizing model of cancer point mutations and their impact on the functionality of the SMC5/6 complex is shown in Supplementary Figure 2.S7.

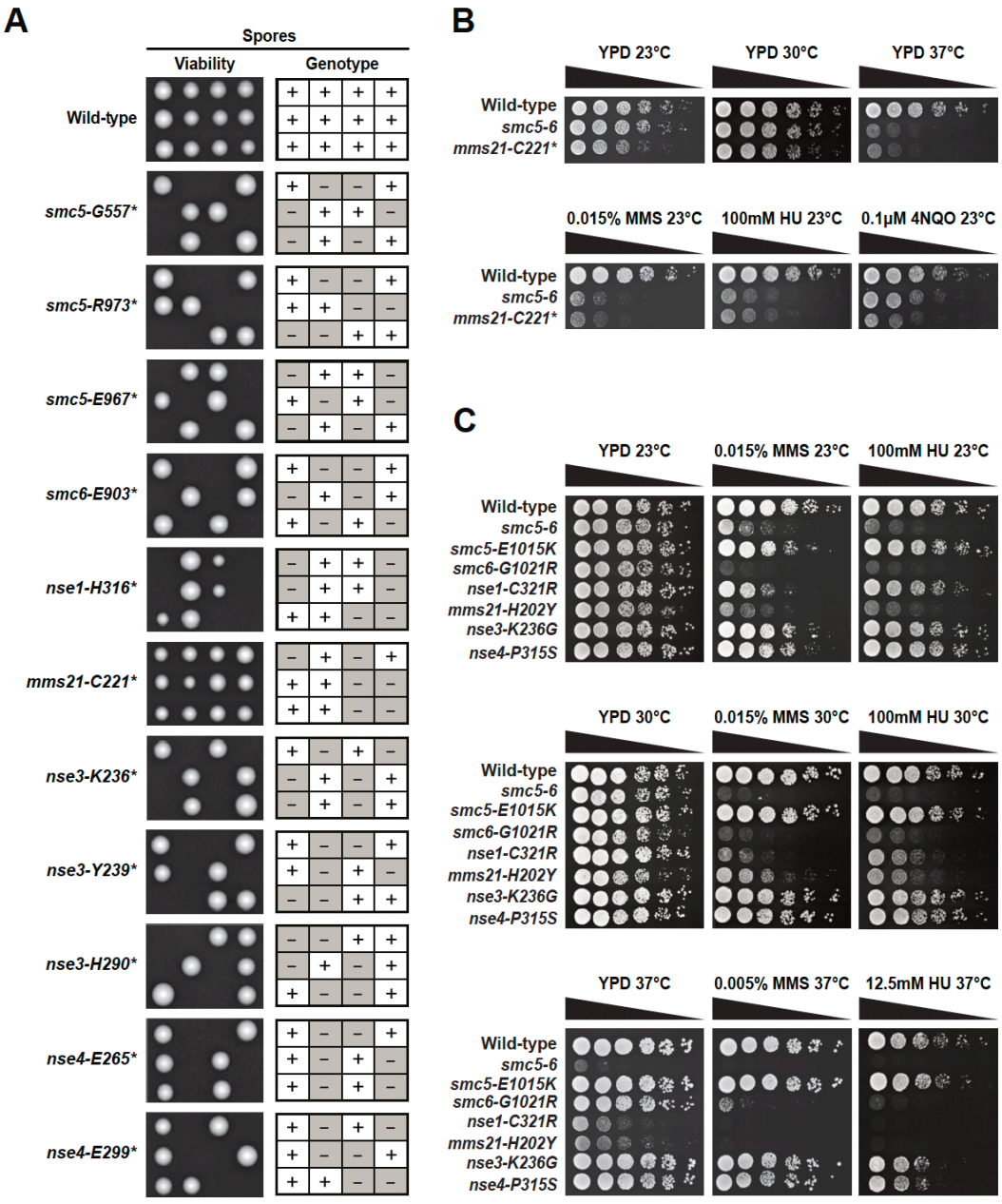


Figure 2. 7: Modeling the phenotype of SMC5/6 cancer mutations in *Saccharomyces cerevisiae*. **(A)** The physiological impact of truncating mutations in SMC5/6 complex components was analyzed by sporulation and dissection of heterozygous diploid yeasts carrying specific mutations in SMC5/6 complex subunits. **(B)** Impact of the SUMO ligase truncation mutation *–mms21-C221*–* in the yeast homolog of human NSMCE2. The proliferation capacity of yeast strains was monitored by dilution assay on solid medium and growth under combinations of genotoxic and high temperature conditions. **(C)** Functional impact of point mutations in the subunits of the SMC5/6 complex. Mutant strains were analyzed as in panel (B).

Since, we observed that several key genes involved in DNA damage repair, cell cycle and DNA replication –especially RAD54B and MCM4– are distinctively linked to SMC5/6 complex alteration and differentially expressed in the SMC5/6 complex alteration group (Figure 2.5B and Supplementary Figure 2.S5), we wanted to test if the protein abundance of yeast Rdh54 (human RAD54B homolog) and yeast Mcm4 would be affected in strains carrying cancer-like mutations in the SMC5/6 complex (Supplementary Figure 2.S8). Remarkably, mutation in glycine 1021 of Smc6 led to a detectable decrease in protein abundance of Mcm4 especially at non-permissive temperature of 37°C (Supplementary Figure 2.S8A). Moreover, yeast strain carrying the *nse3-K236G* alleles had slight decrease in Rdh54 protein expression at 23°C (Supplementary Figure 2.S8B). Together, these results indicate that cancer-like mutations in the yeast SMC5/6 complex reduce its ability to promote an effective response to DNA damage, a phenotype typically associated with reduced genome stability in eukaryotes.

Cancer mutations in SMC5/6 components are haploinsufficient for resistance to replication stress

The majority of the point mutations we identified in SMC5/6 complex components are present in only one allele of the corresponding gene pair found in cancer samples (Figure 2.3A). If they are to have a significant effect on cancer development, one would expect they can act in a dominant manner *in vivo* or, alternatively, that loss of a single allele of a given SMC5/6 component results in haploinsufficient phenotypes (61). To investigate these possibilities, we assessed the viability and growth properties of diploid yeast strains carrying wild-type and cancer

VUS mutations in a heterozygous context. Most of the yeast strains co-expressing wild-type and VUS alleles of SMC5/6 components showed marked growth defects when exposed to various forms of stress (Figure 2.8). For instance, all nonsense mutant alleles we tested in heterozygous diploid conditions showed significant growth defects at 37°C. These defects were substantially exacerbated in the presence of replication stress (Figure 2.8, middle and bottom HU panels). Interestingly, the defective growth behavior of strains carrying heterozygous truncation mutations was similar to that of strains carrying either a full deletion (*smc5Δ / SMC5*) or a previously identified conditional allele (*smc5-6 / SMC5*) of *SMC5*. These results strongly suggest that expression levels of SMC5/6 complex components must be tightly regulated *in vivo* to allow full functionality of the complex, and that loss of 50% protein is sufficient to induce haploinsufficient growth defects in heterozygous strains. Consistent with this, we observed similar growth defects in cells expressing cancer missense mutations in SMC5/6 complex components (Figure 2.8, Top panel). These results dovetail nicely with the morphological haploinsufficiency phenotype observed in strains partly defective in the SMC5/6 complex (62). Taken together, these observations suggest that inactivation of a single allele of SMC5/6 complex components in cancer patients is sufficient to impair the activity of the complex and result in sensitivity to replicative stress.

We note that a cancer mutation identified in the *NSMCE1* gene behaved as a dominant mutation when introduced in yeast. Indeed, a heterozygous diploid strain expressing both the wild-type and *nse1-H316** truncation allele (*i.e.*, corresponding to human *nsmce1-S222**) showed a dominant temperature-sensitive growth defect at 37°C (Figure 2.8, middle panel). The H316 position falls within an important Zinc finger domain in the Nse1 protein that can affect important

functional activities like DNA binding, protein-protein interactions, and various other cellular functions (63). Interestingly, the corresponding amino acid residue in humans (Ser222 of NSMCE1) was found to be mutated in three individual cancer patients. This indicates that a subset of cancer mutations in SMC5/6 complex components may act in a dominant manner to inactivate the complex and promote tumorigenesis.

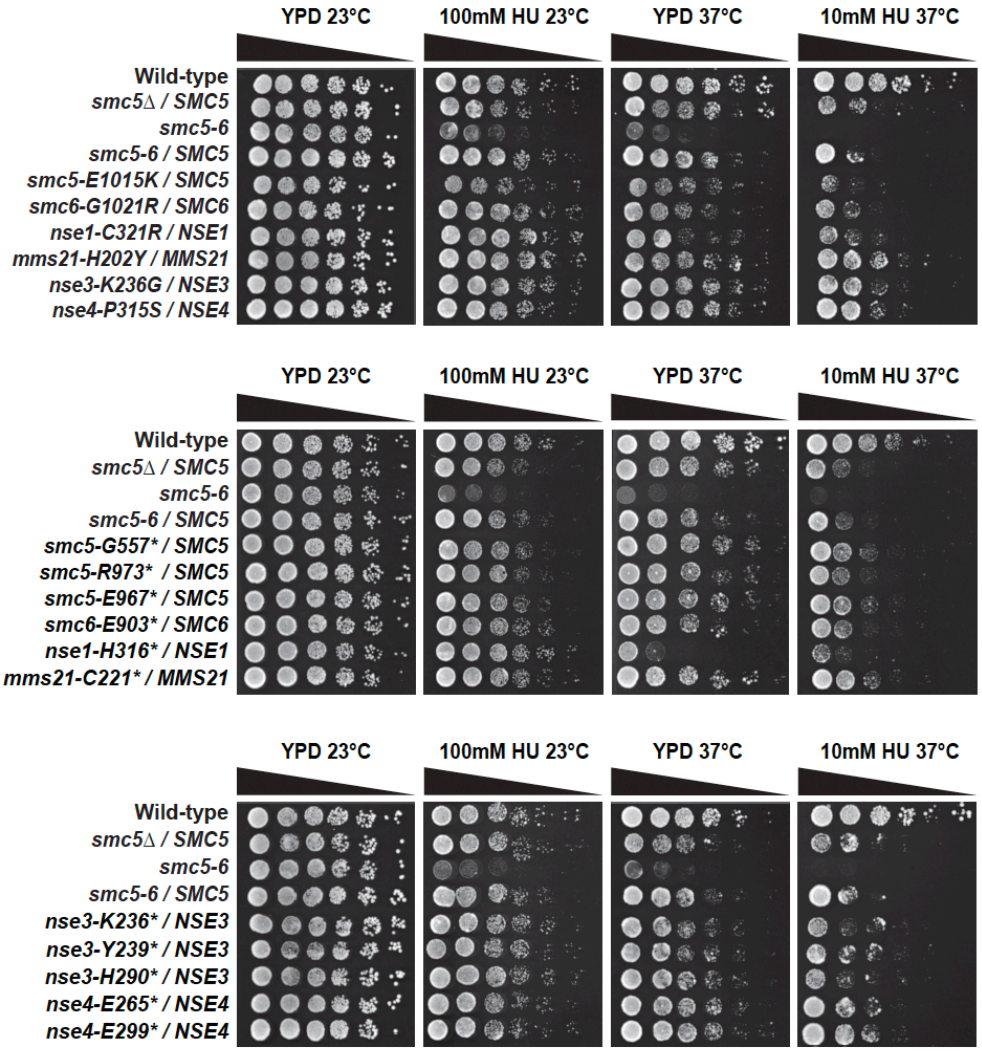


Figure 2. 8: Haploinsufficent phenotype of SMC5/6 cancer mutations in budding yeast. The impact of specific cancer mutations in the subunits of the SMC5/6 complex was assessed in heterozygous diploid yeasts. The proliferation capacity of yeast strains was monitored by dilution assay on solid medium in the presence of genotoxic agents and/or temperature stress, as described in the legend.

2.5 DISCUSSION

Loss of genome stability is a pivotal event that turns healthy cells towards cancerous growth. Whereas the cancer-promoting roles of mutations affecting DNA repair and checkpoint pathways are well established (5, 6), the impact of chromosome-level defects on tumor development is much less understood. In particular, how defects in chromosome morphogenesis (*i.e.*, individualization, compaction and morphology) during mitosis might drive oncogenic processes remain to be fully elucidated. The difficulty in deciphering the contribution of chromosome-level defects in cancer development stems in part from the fact that loss of global chromosome regulation during mitosis is typically associated with cell death, thus impeding detailed functional analysis (7–10). Moreover, when partial loss of global chromosome regulation is observed in human diseases –such as in individuals suffering from hypomorphic mutations in SMC complexes– the associated pathological consequences often lead to lethality at an age that precedes typical timelines of cancer development (64–66). These limitations have considerably hampered progress in understanding how global defects in chromosome morphogenesis might impact cancer development in humans.

To address the challenge outlined above, we took advantage of mutation-rich datasets provided by large-scale efforts to sequence human cancer genomes (32) to understand how alterations in the genes encoding global chromosome regulators might impact oncogenesis. We focused our analysis on the SMC5/6 complex since this is one of the least well understood effector of chromosome architecture in eukaryotes (8, 9). Remarkably, our analysis revealed that components of the SMC5/6 complex are frequently altered in several types of human cancers, including breast (18.05%) and prostate cancers (11.14%). Our data also shows that cancers

carrying co-mutations in multiple subunits of SMC5/6 complex are rare (1.6%) and alterations in single SMC5/6 complex genes are not associated with patient age and sex.

Remarkably, our results establish that alterations in genes encoding the components of the SMC5/6 complex lead to loss of genome stability (*i.e*., higher ploidy score) and adverse patient survival outcome while demonstrating no detectable co-mutational patterns with a panel of genes frequently connected to genome instability. These results are consistent with a recent report by Grange *et al.* (66) showing that a pathogenic variant of the SMC5/6 complex associates with hyperploidy during embryonic development. Our analysis suggests that alterations in the SMC5/6 complex causes hyperploidy in both breast and prostate cancers. DNA ploidy is known to be prognostic to patient survival and a high ploidy score is an indicator of worse prognosis (67, 68). Collectively, these results indicate that the alteration in SMC5/6 complex causes global changes in the stability of cancer genomes, leading to an aggressive disease phenotype and poor patient prognosis.

Our findings also indicate that one of the most frequently affected components of the SMC5/6 complex in human cancer is NSMCE2, a SP-RING domain protein with E3 SUMO ligase activity (14, 15). Previous experiments have shown that this activity is important for DNA repair reactions (14, 15, 69) and our modeling experiments in budding yeast show that cancer point mutations in the SP-RING domain of Mms21 interfere with the maintenance of genome stability after DNA damage. Interestingly, it has been shown that removal of *NSMCE2* in human osteosarcoma (U2OS cells) and human breast cancer (ER^+/PR^+: MCF-7 cells) resulted in significant differences in phenotypes, including slower cell growth and cell cycle arrest (70, 71). Additionally, increased telomerase activity associated with heightened cell survival through activation of ALT

pathways, is detected in 90% of prostate carcinomas (72) and can stem from the amplification of *NSMCE2* gene. In alignment with this finding, we also observe ubiquitous amplification of *NSMCE2* in a large fraction of prostate and breast cancers reported in our study. It is noteworthy that the genomic locus for *NSMCE2* is located near the *MYC* locus on chromosome 8. We have noticed that *NSMCE2* gene is often co-amplified along with the *MYC* gene in an amplicon containing region 8q24 in cancer patients. While it is clear that an increase in *MYC* abundance can act as a driver for cancer formation, it's amplification on double minutes (dmin) does not always lead to *MYC* overexpression (73) suggesting that other genes in the 8q24 amplicon can contribute to oncogenesis. When *NSMCE2* is co-amplified with *MYC* in the same cell, we envision that this would create a highly potent "double hit" scenario because of the synergy associated with simultaneous stimulation of MYC's oncogenic properties and loss of genome stability (due to SMC5/6 complex inactivation). Our results provide primary evidence for this hypothesis. In the MYC-amplified group, NSMCE2 alteration was associated with poor patient survival, demonstrating that simultaneous NSMCE2 and MYC modification is associated with a bleak prognosis (Supplementary Figure 2.S3A, log-rank p-value=0.03). This result further demonstrates the influence of NSMCE2 alterations on patient survival, independent of MYC alterations. Considering the important role of genome instability as a hallmark of cancer (74), it seems highly probable that misregulation of the SMC5/6 complex can promote cancer development as either an independent driver of the process or a cooperator with other oncogenes (such as *MYC*).

Comprehensive analysis of genomic data from cancer patient databases revealed that point mutations and partial deletions occur frequently in all the subunits of the SMC5/6 complex. Prior to our analysis, the functional and tumorigenic impact of these genetic alterations was

unknown, hence the "variant of unknown significance" (VUS) status of SMC5/6 mutations in cancer databases. To shed some light on their functional impact and potential role in cancer development, we introduced several of these cancer-specific mutations at their homologous positions in the subunits of SMC5/6 complex in yeast. Remarkably, all but one SMC5/6 subunit truncations led to lethality in yeast after sporulation and dissection of heterozygous diploid strains. Only the deletion of the C-terminus of yeast Mms21 (human NSMCE2 homolog) led to a viable but sick yeast strain suggesting that even a modest loss of amino-acid sequence can result in a strong impairment of SMC5/6 complex function (14).

Our genetic analysis of diploid yeast strains carrying both wild-type and cancer-specific mutations support the view that expression of a single mutant allele of the SMC5/6 complex is sufficient to weaken the DNA damage response of yeast. Consistent with this, reducing the abundance of Smc5 by half led to a detectable proliferation defect in the *smc5Δ/SMC5* heterozygous mutant, suggesting that mild imbalances in the expression of SMC5/6 complex components can have an impact on the functionality of the complex (75). Similar results have been observed in plants overexpressing *HPY2* (*Arabidopsis thaliana MMS21/NSMCE2*) (76, 77). It is noteworthy that misregulation (including overexpression) of *HPY2* directly impacts ploidy control in plants (hence the name "*High Ploidy2*" for mutants of the *HPY2* gene), an important aspect of genome stability in all eukaryotes (77). This observation dovetails nicely with the hyperploidy phenotype reported by Grange and colleagues in Atelís syndrome patients carrying mutations in the components of the SMC5/6 complex (66). Overall, our data suggest that dual allele inactivation of SMC5/6 complex components is not necessary to promote genome instability and tumorigenesis in patients. Haploinsufficient mutations in genes regulating the DNA

damage response, such as *ATM* and *BLM*, are well known to drive cancer formation (78, 79), which provides a compelling paradigm to directly implicate single-allele mutations of the SMC5/6 complex in cancer development. Consistent with this view, haplo-insufficiency of *NSMCE2* was recently shown to be associated with higher incidence of tumor formation and overall poor survival in a murine model of cancer (26). Taken together, this data supports the view that clinically relevant single-allele mutations affecting evolutionary conserved domains of the SMC5/6 complex or overexpression of some of its components likely contribute as driver mutations for cancer initiation and/or progression.

How might defects in the SMC5/6 complex contribute to cancer initiation and/or maintenance? Hints of the possible answer to this question come from our observation that SMC5/6-defective cancers show dramatically altered ploidy levels. It is now widely appreciated that changes in the chromosome contents of cells, in particular aneuploidy (1–4), create a substantial burden on many cellular processes and can promote cancer development. For instance, high level of aneuploidy is a major driving force for prostate cancer development, is indicative of a higher extent of aggressiveness in primary prostate cancers (80) and can confer resistance to chemotherapy (81, 82). Our work revealed that patients with altered SMC5/6, particularly with altered *NSMCE2*, have high aneuploidy scores in prostate and breast cancers, which outlines a likely mechanism by which SMC5/6 complex mutations can promote cancer development. Interestingly, defects in SMC5/6 activity might also provide a unique window of opportunity for cancer treatment. Extensive analysis of genetic interaction networks in yeast have identified several synthetic lethal interactions connecting members of the SMC5/6 complex with proteins involved in the resolution of DNA replication damage. Many of these proteins have

clear homologs in higher eukaryotes and are known to be mutated in several human cancers (26, 83, 84). Since many genetic interactions are conserved in eukaryotes, we propose that synthetic lethal partners of the SMC5/6 complex –particularly those involved in the repair of DNA replication damage– are likely to be effective targets for therapeutic intervention in cancers harboring mutations in the SMC5/6 complex. This prediction is further supported by our observation that yeast strains carrying cancer-specific mutations in its SMC5/6 complex are highly sensitive to replication-stress induced by HU. Since HU and its derivatives have been shown to be effective anti-cancer drugs with minimal side effects (85–88), we predict cancers harboring mutations in SMC5/6 complex components would respond well to this drug or other types of DNA replication stressors (*e.g.*, CHK1 inhibitors) (88).

In conclusion, our work demonstrates the pervasiveness of mutations affecting the SMC5/6 complex in diverse human cancers and strongly suggests that patients harboring SMC5/6-defective cancers would benefit from precision/genotype-targeted treatment modalities focused on creating synthetic lethality connected to DNA replication stress. Future efforts will be required to faithfully manipulate specific functional pathways connected with the SMC5/6 complex and leverage that knowledge to develop better diagnostic procedures and therapeutic agent to treat human cancers.

DATA AND CODE AVAILABILITY

Patient data including mutation, DNA ploidy score and survival information used in this study are available at the cBioPortal for Cancer Genomics (https://cbioportal.org) web resources. METABRIC (36) study gene expression data is available through MetaGxBreast package (version

1.18.0) (37). Processed next generation sequencing based gene expression data is available through TCGAbiolinks package (version 2.25.3) (41). All necessary code to reproduce the bioinformatic analyses and figures described in this manuscript are available on the Zenodo open repository site at https://doi.org/10.5281/zenodo.8256896. All accompanying data can be found on Zenodo at https://doi.org/10.5281/zenodo.8256823.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Cancer Online.

AUTHOR CONTRIBUTIONS

SR and DD conceived and designed yeast experiments; AZ and AM performed all bioinformatics analyses; SR created yeast strains, assessed their viability and DNA damage sensitivity; SR and AZ, prepared figures; AZ and AM wrote all the bioinformatics sections of the paper; SR and DD wrote the sections of the paper pertaining to yeast experiments.

ACKNOWLEDGMENTS

We thank members of the D'Amours laboratory for their comments on the manuscript, and Dr. Hemanta Adhikary in particular for his help with the structural model of the SMC5/6 complex. This work was supported by grants to DD (CIHR FDN–167265) and AM (J.P. Bickell Foundation Medical Research Grant). DD is supported by a Canada Research Chair in Chromatin Dynamics & Genome Architecture.

DECLARATION OF INTERESTS

The authors declare no competing interests.

2.6 REFERENCES

1. Keuper,K., Wieland,A., Räschle,M. and Storchova,Z. (2021) Processes shaping cancer genomes - From mitotic defects to chromosomal rearrangements. *DNA Repair (Amst)*, **107**, 103207.

2. Kneissig,M., Bernhard,S. and Storchova,Z. (2019) Modelling chromosome structural and copy number changes to understand cancer genomes. *Curr Opin Genet Dev*, **54**, 25–32.

3. Garribba,L. and Santaguida,S. (2022) The Dynamic Instability of the Aneuploid Genome. *Front Cell Dev Biol*, **10**, 838928.

4. Ben-David,U. and Amon,A. (2020) Context is everything: aneuploidy in cancer. *Nat Rev Genet*, **21**, 44–62.

5. Groelly,F.J., Fawkes,M., Dagg,R.A., Blackford,A.N. and Tarsounas,M. (2023) Targeting DNA damage response pathways in cancer. *Nat Rev Cancer*, **23**, 78–94.

6. Klinakis,A., Karagiannis,D. and Rampias,T. (2020) Targeting DNA repair in cancer: current state and novel approaches. *Cell. Mol. Life Sci.*, **77**, 677–703.

7. Yatskevich,S., Rhodes,J. and Nasmyth,K. (2019) Organization of Chromosomal DNA by SMC Complexes. *Annu. Rev. Genet.*, **53**, 445–482.

8. Aragón,L. (2018) The Smc5/6 Complex: New and Old Functions of the Enigmatic Long-Distance Relative. *Annu. Rev. Genet.*, **52**, 89–107.

9. Solé-Soler,R. and Torres-Rosell,J. (2020) Smc5/6, an atypical SMC complex with two RING-type subunits. *Biochem Soc Trans*, **48**, 2159–2171.

10. Kim,Y. and Yu,H. (2020) Shaping of the 3D genome by the ATPase machine cohesin. *Exp Mol Med*, **52**, 1891–1897.

11. De Koninck,M. and Losada,A. (2016) Cohesin Mutations in Cancer. *Cold Spring Harb Perspect Med*, **6**, a026476.

12. Yuen,K.C. and Gerton,J.L. (2018) Taking cohesin and condensin in context. *PLOS Genetics*, **14**, e1007118.

13. Leiserson,M.D.M., Vandin,F., Wu,H.-T., Dobson,J.R., Eldridge,J.V., Thomas,J.L., Papoutsaki,A., Kim,Y., Niu,B., McLellan,M., *et al.* (2015) Pan-cancer network analysis identifies

combinations of rare somatic mutations across pathways and protein complexes. *Nat Genet*, **47**, 106–114.

14. Zhao,X. and Blobel,G. (2005) A SUMO ligase is part of a nuclear multiprotein complex that affects DNA repair and chromosomal organization. *Proc Natl Acad Sci U S A*, **102**, 4777–4782.

15. Potts,P.R. and Yu,H. (2005) Human MMS21/NSE2 is a SUMO ligase required for DNA repair. *Mol Cell Biol*, **25**, 7021–7032.

16. Dhingra,N. and Zhao,X. (2021) Advances in SUMO-based regulation of homologous recombination. *Curr Opin Genet Dev*, **71**, 114–119.

17. Potts,P.R. and Yu,H. (2007) The SMC5/6 complex maintains telomere length in ALT cancer cells through SUMOylation of telomere-binding proteins. *Nature Structural &Amp; Molecular Biology*, **14**, 581.

18. Noël,J.-F. and Wellinger,R.J. (2011) Abrupt telomere losses and reduced end-resection can explain accelerated senescence of Smc5/6 mutants lacking telomerase. *DNA Repair (Amst)*, **10**, 271–282.

19. Taschner,M., Basquin,J., Steigenberger,B., Schäfer,I.B., Soh,Y.-M., Basquin,C., Lorentzen,E., Räschle,M., Scheltema,R.A. and Gruber,S. (2021) Nse5/6 inhibits the Smc5/6 ATPase and modulates DNA substrate binding. *EMBO J*, **40**, e107807.

20. Yu,Y., Li,S., Ser,Z., Sanyal,T., Choi,K., Wan,B., Kuang,H., Sali,A., Kentsis,A., Patel,D.J., *et al.* (2021) Integrative analysis reveals unique structural and functional features of the Smc5/6 complex. *Proc Natl Acad Sci U S A*, **118**, e2026844118.

21. Yu,Y., Li,S., Ser,Z., Kuang,H., Than,T., Guan,D., Zhao,X. and Patel,D.J. (2022) Cryo-EM structure of DNA-bound Smc5/6 reveals DNA clamping enabled by multi-subunit conformational changes. *Proc Natl Acad Sci U S A*, **119**, e2202799119.

22. Hallett,S.T., Schellenberger,P., Zhou,L., Beuron,F., Morris,E., Murray,J.M. and Oliver,A.W. (2021) Nse5/6 is a negative regulator of the ATPase activity of the Smc5/6 complex. *Nucleic Acids Res*, **49**, 4534–4549.

23. Adamus,M., Lelkes,E., Potesil,D., Ganji,S.R., Kolesar,P., Zabrady,K., Zdrahal,Z. and Palecek,J.J. (2020) Molecular Insights into the Architecture of the Human SMC5/6 Complex. *J Mol Biol*, **432**, 3820–3837.

24. Serrano,D., Cordero,G., Kawamura,R., Sverzhinsky,A., Sarker,M., Roy,S., Malo,C., Pascal,J.M., Marko,J.F. and D'Amours,D. (2020) The Smc5/6 Core Complex Is a Structure-Specific DNA Binding and Compacting Machine. *Mol Cell*, **80**, 1025-1038.e5.

25. Gutierrez-Escribano,P., Hormeño,S., Madariaga-Marcos,J., Solé-Soler,R., O'Reilly,F.J., Morris,K., Aicart-Ramos,C., Aramayo,R., Montoya,A., Kramer,H., *et al.* (2020) Purified Smc5/6 Complex Exhibits DNA Substrate Recognition and Compaction. *Mol Cell*, **80**, 1039-1054.e6.

26. Jacome,A., Gutierrez-Martinez,P., Schiavoni,F., Tenaglia,E., Martinez,P., Rodríguez-Acebes,S., Lecona,E., Murga,M., Méndez,J., Blasco,M.A., *et al.* (2015) NSMCE2 suppresses cancer and aging in mice independently of its SUMO ligase activity. *The EMBO Journal*, **34**, 2604–2619.

27. Nie,H., Wang,Y., Yang,X., Liao,Z., He,X., Zhou,J. and Ou,C. (2021) Clinical Significance and Integrative Analysis of the SMC Family in Hepatocellular Carcinoma. *Front Med (Lausanne)*, **8**, 727965.

28. Yang,H., Gao,S., Chen,J. and Lou,W. (2020) UBE2I promotes metastasis and correlates with poor prognosis in hepatocellular carcinoma. *Cancer Cell Int*, **20**, 234.

29. Zhou,J., Wu,G., Tong,Z., Sun,J., Su,J., Cao,Z., Luo,Y. and Wang,W. (2020) Prognostic relevance of SMC family gene expression in human sarcoma. *Aging (Albany NY)*, **13**, 1473–1487.

30. Di Benedetto,C., Oh,J., Choudhery,Z., Shi,W., Valdes,G. and Betancur,P. (2022) NSMCE2, a novel super-enhancer-regulated gene, is linked to poor prognosis and therapy resistance in breast cancer. *BMC Cancer*, **22**, 1056.

31. Saunus,J.M., Quinn,M.C.J., Patch,A.-M., Pearson,J.V., Bailey,P.J., Nones,K., McCart Reed,A.E., Miller,D., Wilson,P.J., Al-Ejeh,F., *et al.* (2015) Integrated genomic and transcriptomic analysis of human brain metastases identifies alterations of potential clinical significance. *J Pathol*, **237**, 363–378.

32. Gao,J., Aksoy,B.A., Dogrusoz,U., Dresdner,G., Gross,B., Sumer,S.O., Sun,Y., Jacobsen,A., Sinha,R., Larsson,E., *et al.* (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*, **6**, pl1.

33. Colaprico,A., Silva,T.C., Olsen,C., Garofano,L., Cava,C., Garolini,D., Sabedot,T.S., Malta,T.M., Pagnotta,S.M., Castiglioni,I., *et al.* (2016) TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*, **44**, e71.

34. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, **43**, e47.

35. Chen,Y., Lun,A.T.L. and Smyth,G.K. (2016) From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res*, **5**, 1438.

36. Curtis,C., Shah,S.P., Chin,S.-F., Turashvili,G., Rueda,O.M., Dunning,M.J., Speed,D., Lynch,A.G., Samarajiwa,S., Yuan,Y., *et al.* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, **486**, 346–352.

37. Gendoo,D.M.A., Zon,M., Sandhu,V., Manem,V.S.K., Ratanasirigulchai,N., Chen,G.M., Waldron,L. and Haibe-Kains,B. (2019) MetaGxData: Clinically Annotated Breast, Ovarian and Pancreatic Cancer Datasets and their Use in Generating a Multi-Cancer Gene Signature. *Sci Rep*, **9**, 8770.

38. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

39. Cancer Genome Atlas Research Network, Weinstein,J.N., Collisson,E.A., Mills,G.B., Shaw,K.R.M., Ozenberger,B.A., Ellrott,K., Shmulevich,I., Sander,C. and Stuart,J.M. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*, **45**, 1113–1120.

40. Korotkevich,G., Sukhov,V., Budin,N., Shpak,B., Artyomov,M.N. and Sergushichev,A. (2021) Fast gene set enrichment analysis. *bioRxiv*, 10.1101/060012.

41. Yu,G., Wang,L.-G., Yan,G.-R. and He,Q.-Y. (2015) DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics*, **31**, 608–609.

42. Broer,L., Lill,C.M., Schuur,M., Amin,N., Roehr,J.T., Bertram,L., Ioannidis,J.P.A. and van Duijn,C.M. (2013) Distinguishing true from false positives in genomic studies: p values. *Eur J Epidemiol*, **28**, 131–138.

43. St-Pierre,J., Douziech,M., Bazile,F., Pascariu,M., Bonneil,E., Sauvé,V., Ratsima,H. and D'Amours,D. (2009) Polo kinase regulates mitotic chromosome condensation by hyperactivation of condensin DNA supercoiling activity. *Mol Cell*, **34**, 416–426.

44. Sinha,H., David,L., Pascon,R.C., Clauder-Münster,S., Krishnakumar,S., Nguyen,M., Shi,G., Dean,J., Davis,R.W., Oefner,P.J., *et al.* (2008) Sequential Elimination of Major-Effect Contributors Identifies Additional Quantitative Trait Loci Conditioning High-Temperature Growth in Yeast. *Genetics*, **180**, 1661–1670.

45. Auesukaree,C., Damnernsawad,A., Kruatrachue,M., Pokethitiyook,P., Boonchird,C., Kaneko,Y. and Harashima,S. (2009) Genome-wide identification of genes involved in tolerance to various environmental stresses in Saccharomyces cerevisiae. *J Appl Genet*, **50**, 301–310.

46. Ratsima,H., Ladouceur,A.-M., Pascariu,M., Sauvé,V., Salloum,Z., Maddox,P.S. and D'Amours,D. (2011) Independent modulation of the kinase and polo-box activities of Cdc5 protein unravels unique roles in the maintenance of genome stability. *Proc Natl Acad Sci U S A*, **108**, E914-923.

47. Foiani,M., Marini,F., Gamba,D., Lucchini,G. and Plevani,P. (1994) The B subunit of the DNA polymerase alpha-primase complex in Saccharomyces cerevisiae executes an essential function at the initial stage of DNA replication. *Mol Cell Biol*, **14**, 923–933.

48. Duan,X., Sarangi,P., Liu,X., Rangi,G.K., Zhao,X. and Ye,H. (2009) Structural and functional insights into the roles of the Mms21 subunit of the Smc5/6 complex. *Mol Cell*, **35**, 657–668.

49. Varejão,N., Lascorz,J., Codina-Fabra,J., Bellí,G., Borràs-Gas,H., Torres-Rosell,J. and Reverter,D. (2021) Structural basis for the E3 ligase activity enhancement of yeast Nse2 by SUMO-interacting motifs. *Nat Commun*, **12**, 7013.

50. Båvner,A., Matthews,J., Sanyal,S., Gustafsson,J.-A. and Treuter,E. (2005) EID3 is a novel EID family member and an inhibitor of CBP-dependent co-activation. *Nucleic Acids Res*, **33**, 3561–3569.

51. Xu,C., Liu,D., Mei,H., Hu,J. and Luo,M. (2019) Knockdown of RAD54B expression reduces cell proliferation and induces apoptosis in lung cancer cells. *J Int Med Res*, **47**, 5650–5659.

52. Miyagawa,K., Tsuruga,T., Kinomura,A., Usui,K., Katsura,M., Tashiro,S., Mishima,H. and Tanaka,K. (2002) A role for RAD54B in homologous recombination in human cells. *The EMBO Journal*, **21**, 175–180.

53. Wesoly,J., Agarwal,S., Sigurdsson,S., Bussen,W., Van Komen,S., Qin,J., van Steeg,H., van Benthem,J., Wassenaar,E., Baarends,W.M., *et al.* (2006) Differential Contributions of Mammalian Rad54 Paralogs to Recombination, DNA Damage Repair, and Meiosis. *Molecular and Cellular Biology*, **26**, 976–989.

54. Payton,M., Scully,S., Chung,G. and Coats,S. (2002) Deregulation of cyclin E2 expression and associated kinase activity in primary breast tumors. *Oncogene*, **21**, 8529–8534.

55. Caldon,C.E., Sergio,C.M., Sutherland,R.L. and Musgrove,E.A. (2013) Differences in degradation lead to asynchronous expression of cyclin E1 and cyclin E2 in cancer cells. *Cell Cycle*, **12**, 596–605.

56. Lu,H., Shamanna,R.A., Keijzers,G., Anand,R., Rasmussen,L.J., Cejka,P., Croteau,D.L. and Bohr,V.A. (2016) RECQL4 Promotes DNA End Resection in Repair of DNA Double-Strand Breaks. *Cell Rep*, **16**, 161–173.

57. Ishimi,Y., Komamura-Kohno,Y., Kwon,H.-J., Yamada,K. and Nakanishi,M. (2003) Identification of MCM4 as a target of the DNA replication block checkpoint system. *J Biol Chem*, **278**, 24644–24650.

58. Kalev,P., Simicek,M., Vazquez,I., Munck,S., Chen,L., Soin,T., Danda,N., Chen,W. and Sablina,A. (2012) Loss of PPP2R2A inhibits homologous recombination DNA repair and predicts tumor sensitivity to PARP inhibition. *Cancer Res*, **72**, 6414–6424.

59. Squatrito,M., Vanoli,F., Schultz,N., Jasin,M. and Holland,E.C. (2012) 53BP1 is a haploinsufficient tumor suppressor and protects cells from radiation response in glioma. *Cancer Res*, **72**, 5250–5260.

60. Torres-Rosell,J., Machín,F., Farmer,S., Jarmuz,A., Eydmann,T., Dalgaard,J.Z. and Aragón,L. (2005) SMC5 and SMC6 genes are required for the segregation of repetitive chromosome regions. *Nat Cell Biol*, **7**, 412–419.

61. Veitia,R.A. (2002) Exploring the etiology of haploinsufficiency. *Bioessays*, **24**, 175–184.

62. Ohnuki,S. and Ohya,Y. (2018) High-dimensional single-cell phenotyping reveals extensive haploinsufficiency. *PLoS Biol*, **16**, e2005130.

63. Pebernard,S., Perry,J.J.P., Tainer,J.A. and Boddy,M.N. (2008) Nse1 RING-like Domain Supports Functions of the Smc5-Smc6 Holocomplex in Genome Stability. *Mol Biol Cell*, **19**, 4099–4109.

64. van der Crabben,S.N., Hennus,M.P., McGregor,G.A., Ritter,D.I., Nagamani,S.C.S., Wells,O.S., Harakalova,M., Chinn,I.K., Alt,A., Vondrova,L., *et al.* (2016) Destabilized SMC5/6 complex leads to chromosome breakage syndrome with severe lung disease. *Journal of Clinical Investigation*, **126**, 2881–2892.

65. Payne,F., Colnaghi,R., Rocha,N., Seth,A., Harris,J., Carpenter,G., Bottomley,W.E., Wheeler,E., Wong,S., Saudek,V., *et al.* (2014) Hypomorphism in human NSMCE2 linked to primordial dwarfism and insulin resistance. *J. Clin. Invest.*, **124**, 4028–4038.

66. Grange,L.J., Reynolds,J.J., Ullah,F., Isidor,B., Shearer,R.F., Latypova,X., Baxley,R.M., Oliver,A.W., Ganesh,A., Cooke,S.L., *et al.* (2022) Pathogenic variants in SLF2 and SMC5 cause segmented chromosomes and mosaic variegated hyperploidy. *Nat Commun*, **13**, 6664.

67. Pinto,A.E., Pereira,T., Santos,M., Branco,M., Dias,A., Silva,G.L., Ferreira,M.C. and André,S. (2013) DNA ploidy is an independent predictor of survival in breast invasive ductal carcinoma: a long-term multivariate analysis of 393 patients. *Ann Surg Oncol*, **20**, 1530–1537.

68. Martínez-Jabaloyas,J.M., Ruiz-Cerdá,J.L., Hernández,M., Jiménez,A. and Jiménez-Cruz,F. (2002) Prognostic value of DNA ploidy and nuclear morphometry in prostate cancer treated with androgen deprivation. *Urology*, **59**, 715–720.

69. Zapatka,M., Pociño-Merino,I., Heluani-Gahete,H., Bermúdez-López,M., Tarrés,M., Ibars,E., Solé-Soler,R., Gutiérrez-Escribano,P., Apostolova,S., Casas,C., *et al.* (2019) Sumoylation of Smc5 Promotes Error-free Bypass at Damaged Replication Forks. *Cell Reports*, **29**, 3160-3172.e4.

70. Verver,D.E., Zheng,Y., Speijer,D., Hoebe,R., Dekker,H.L., Repping,S., Stap,J. and Hamer,G. (2016) Non-SMC Element 2 (NSMCE2) of the SMC5/6 Complex Helps to Resolve Topological Stress. *International Journal of Molecular Sciences*, **17**, 1782.

71. Ni,H.-J., Chang,Y.-N., Kao,P.-H., Chai,S.-P., Hsieh,Y.-H., Wang,D.-H. and Fong,J.C. (2012) Depletion of SUMO ligase hMMS21 impairs G1 to S transition in MCF-7 breast cancer cells. *Biochimica et Biophysica Acta (BBA) - General Subjects*, **1820**, 1893–1900.

72. Graham,M.K. and Meeker,A. (2017) Telomeres and telomerase in prostate cancer development and therapy. *Nat Rev Urol*, **14**, 607–619.

73. Storlazzi,C.T., Fioretos,T., Paulsson,K., Strömbeck,B., Lassen,C., Ahlgren,T., Juliusson,G., Mitelman,F., Rocchi,M. and Johansson,B. (2004) Identification of a commonly amplified 4.3 Mb region with overexpression of C8FW, but not MYC in MYC-containing double minutes in myeloid malignancies. *Hum Mol Genet*, **13**, 1479–1485.

74. Hanahan,D. and Weinberg,R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.

75. Papp,B., Pál,C. and Hurst,L.D. (2003) Dosage sensitivity and the evolution of gene families in yeast. *Nature*, **424**, 194–197.

76. Zhang,S., Qi,Y., Liu,M. and Yang,C. (2013) SUMO E3 ligase AtMMS21 regulates drought tolerance in Arabidopsis thaliana(F). *J Integr Plant Biol*, **55**, 83–95.

77. Lee,K.-J., Kim,Y.-E., Lee,H. and Park,S.-Y. (2017) Overexpression of SUMO E3 ligase HPY2 regulates the cell cycle in petunia development. *Hortic. Environ. Biotechnol.*, **58**, 384–392.

78. Spring,K., Ahangari,F., Scott,S.P., Waring,P., Purdie,D.M., Chen,P.C., Hourigan,K., Ramsay,J., McKinnon,P.J., Swift,M., *et al.* (2002) Mice heterozygous for mutation in Atm, the gene involved in ataxia-telangiectasia, have heightened susceptibility to cancer. *Nat Genet*, **32**, 185–190.

79. Goss,K.H., Risinger,M.A., Kordich,J.J., Sanz,M.M., Straughen,J.E., Slovek,L.E., Capobianco,A.J., German,J., Boivin,G.P. and Groden,J. (2002) Enhanced tumor formation in mice heterozygous for Blm mutation. *Science*, **297**, 2051–2053.

80. Stopsack,K.H., Whittaker,C.A., Gerke,T.A., Loda,M., Kantoff,P.W., Mucci,L.A. and Amon,A. (2019) Aneuploidy drives lethal progression in prostate cancer. *Proceedings of the National Academy of Sciences*, **116**, 11390–11395.

81. Ippolito,M.R., Martis,V., Martin,S., Tijhuis,A.E., Hong,C., Wardenaar,R., Dumont,M., Zerbib,J., Spierings,D.C.J., Fachinetti,D., *et al.* (2021) Gene copy-number changes and chromosomal instability induced by aneuploidy confer resistance to chemotherapy. *Dev Cell*, **56**, 2440-2454.e6.

82. Lukow,D.A., Sausville,E.L., Suri,P., Chunduri,N.K., Wieland,A., Leu,J., Smith,J.C., Girish,V., Kumar,A.A., Kendall,J., *et al.* (2021) Chromosomal instability accelerates the evolution of resistance to anti-cancer therapies. *Dev Cell*, **56**, 2427-2439.e4.

83. Lafuente-Barquero,J., Luke-Glaser,S., Graf,M., Silva,S., Gómez-González,B., Lockhart,A., Lisby,M., Aguilera,A. and Luke,B. (2017) The Smc5/6 complex regulates the yeast Mph1 helicase at RNA-DNA hybrid-mediated DNA damage. *PLOS Genetics*, **13**, e1007136.

84. Agashe,S., Joseph,C.R., Reyes,T.A.C., Menolfi,D., Giannattasio,M., Waizenegger,A., Szakal,B. and Branzei,D. (2021) Smc5/6 functions with Sgs1-Top3-Rmi1 to complete chromosome replication at natural pause sites. *Nat Commun*, **12**, 2111.

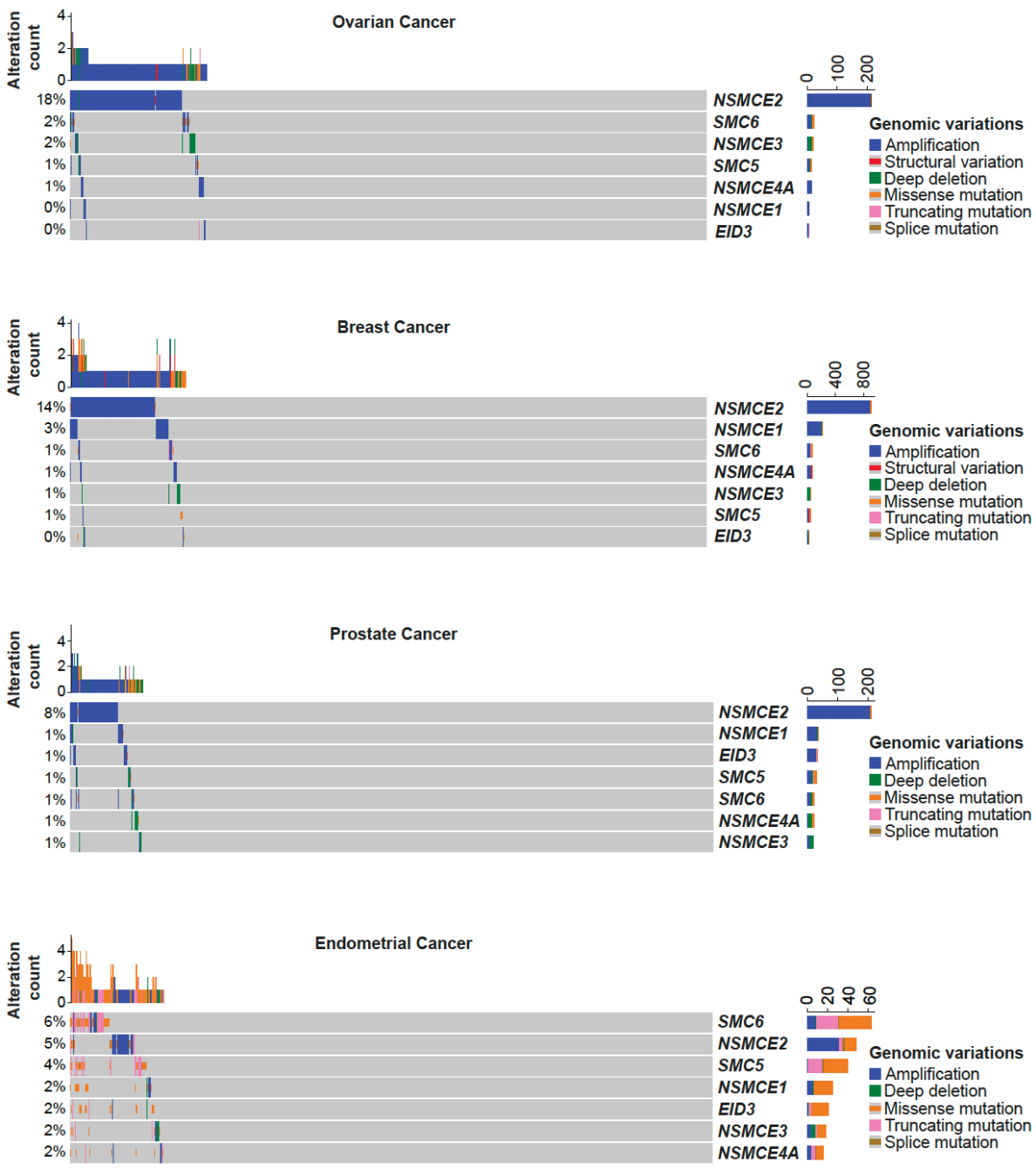
85. Madaan,K., Kaushik,D. and Verma,T. (2012) Hydroxyurea: a key player in cancer chemotherapy. *Expert Review of Anticancer Therapy*, **12**, 19–29.

86. Saban,N. and Bujak,M. (2009) Hydroxyurea and hydroxamic acid derivatives as antitumor drugs. *Cancer Chemother Pharmacol*, **64**, 213–221.

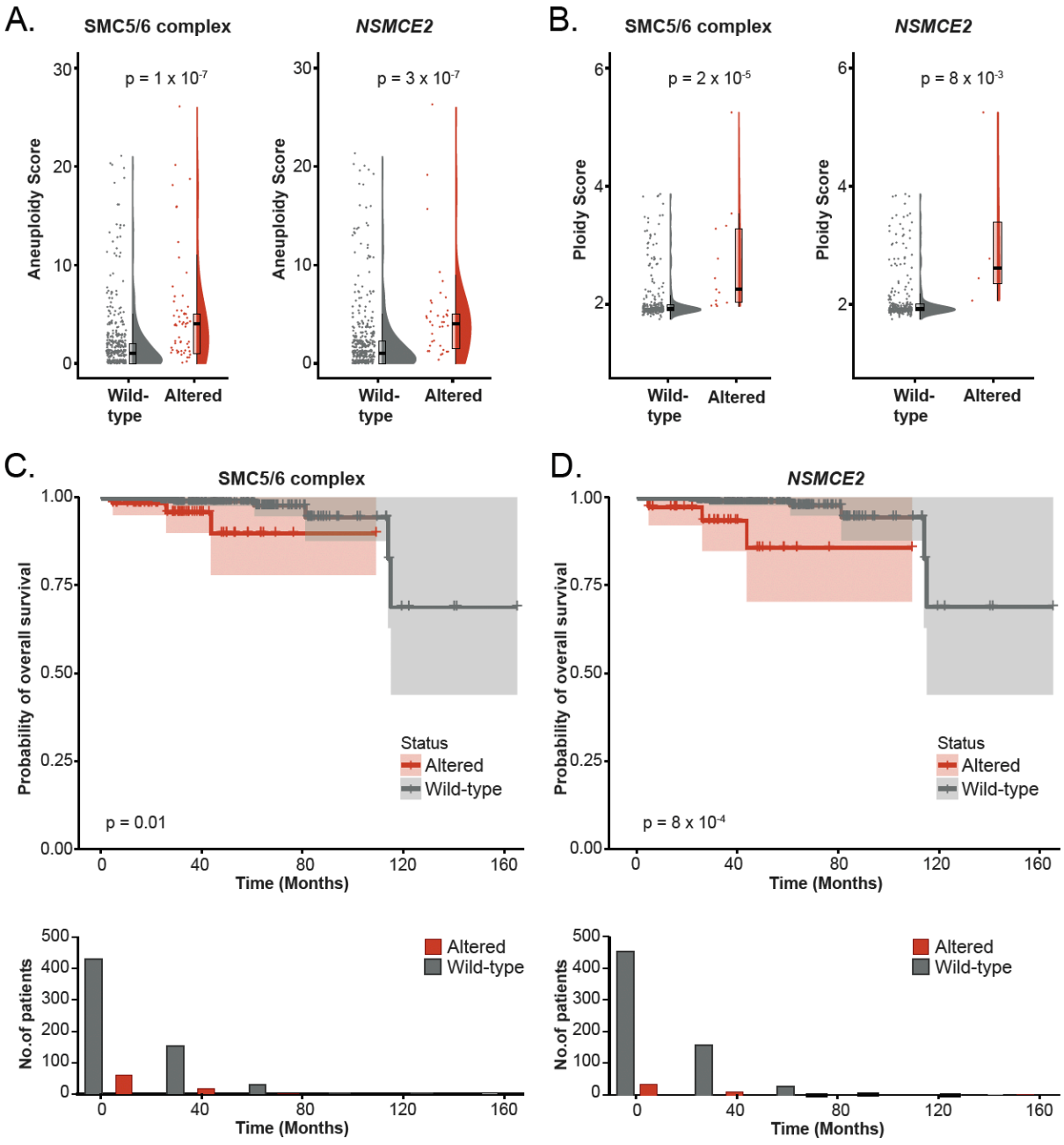
87. Nazareth,D., Jones,M.J.K. and Gabrielli,B. (2019) Everything in Moderation: Lessons Learned by Exploiting Moderate Replication Stress in Cancer. *Cancers*, **11**, 1320.

88. Oo,Z.Y., Proctor,M., Stevenson,A.J., Nazareth,D., Fernando,M., Daignault,S.M., Lanagan,C., Walpole,S., Bonazzi,V., Škalamera,D., *et al.* (2019) Combined use of subclinical hydroxyurea and CHK1 inhibitor effectively controls melanoma and lung cancer progression, with reduced normal tissue toxicity compared to gemcitabine. *Molecular Oncology*, **13**, 1503–1518.

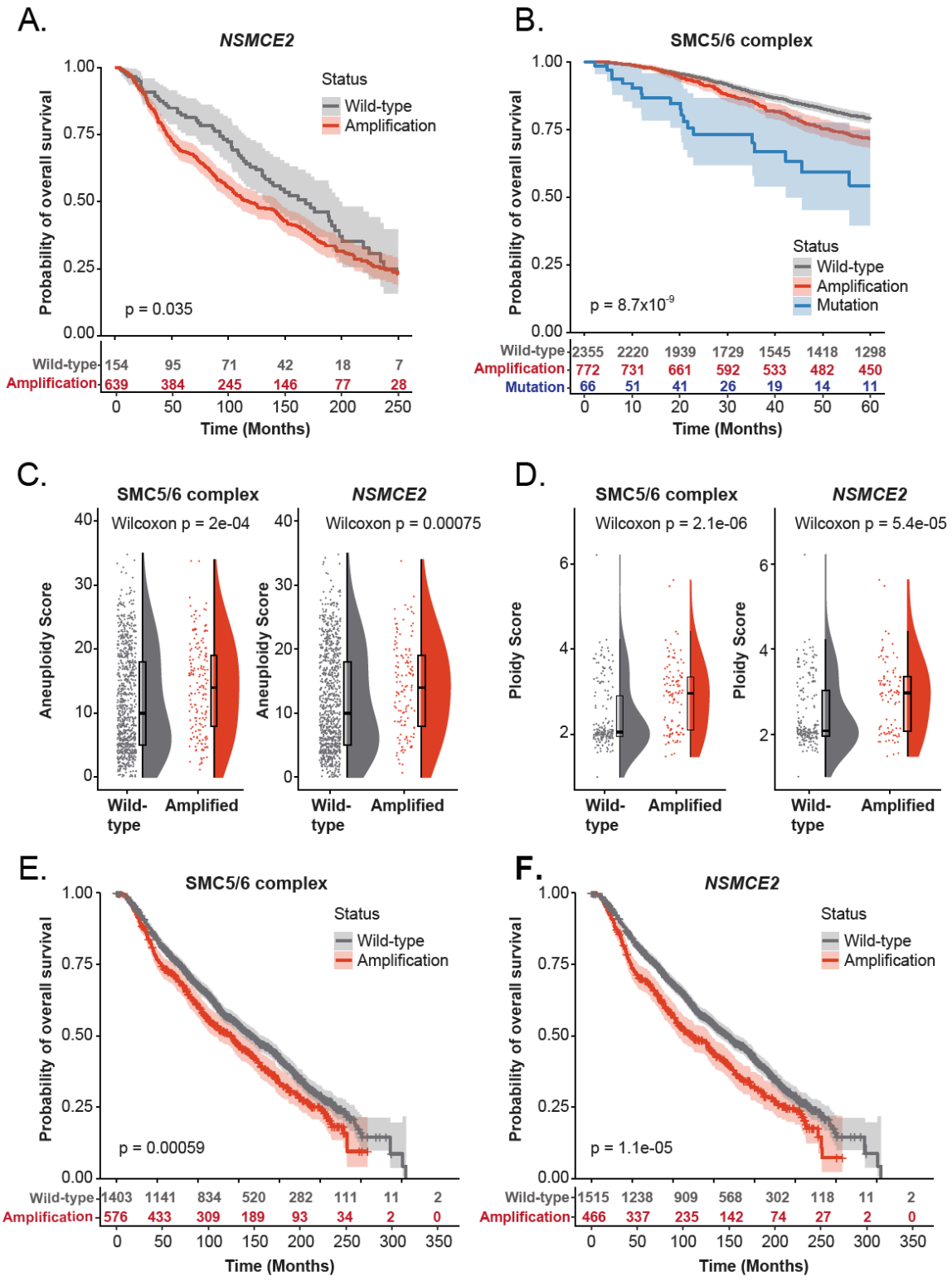
2.7 SUPPLEMENTARY INFORMATION



Supplementary Figure 2.S 1: Oncoprints visualization of genetic aberration in the SMC5/6 complex genes in ovarian, breast, prostate and endometrial cancer.

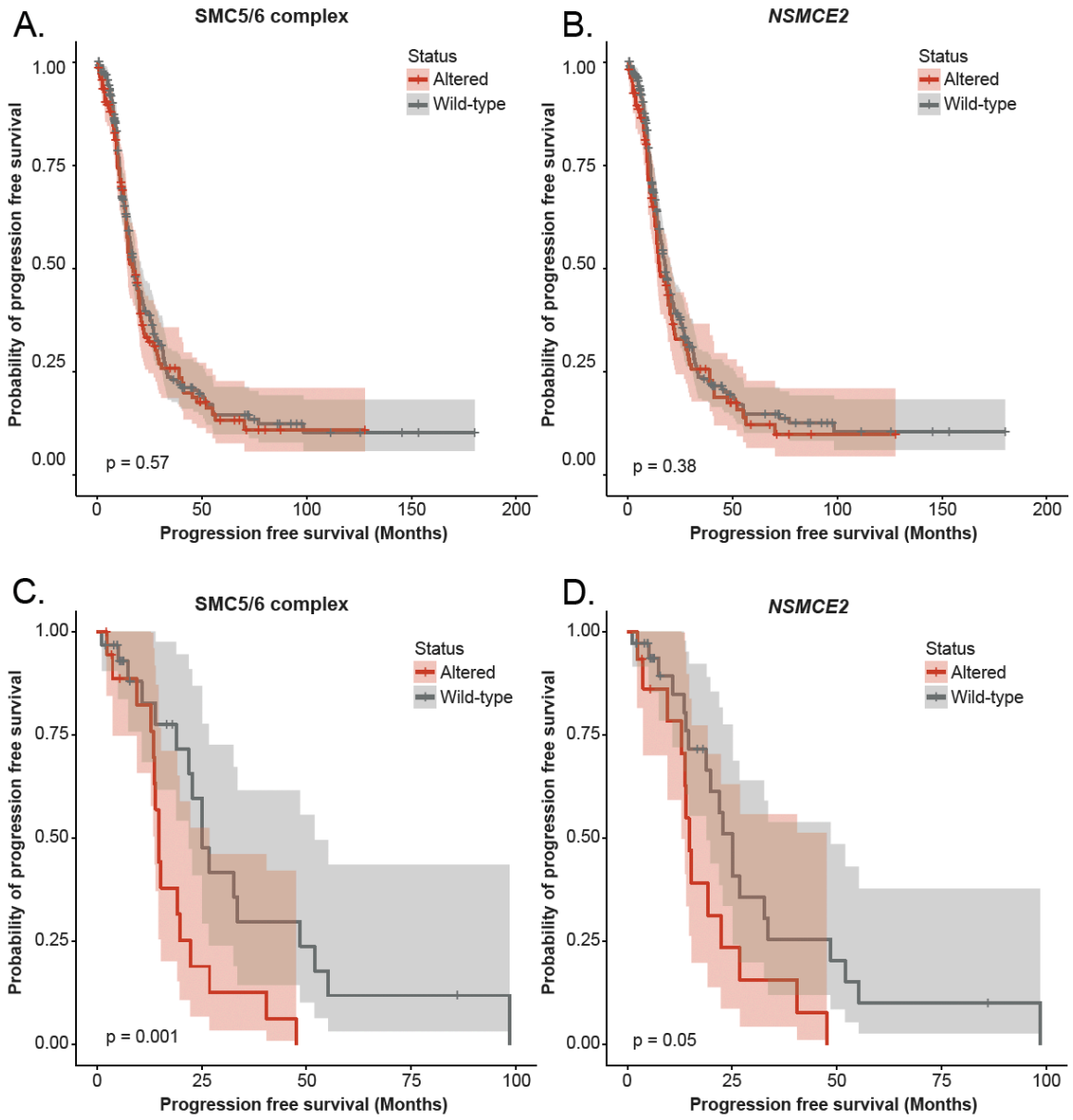


Supplementary Figure 2.S 2:Distribution of ploidy score, and survival analysis in prostate cancer. Distribution of aneuploidy score (A) and ploidy score (B) in the TCGA prostate cohort. Survival analysis of TCGA prostate cohort with or without genomic alteration in the SMC5/6 complex (C) and NSMCE2 (D). Risk tables are represented as bar charts in the bottom of the KM plot.

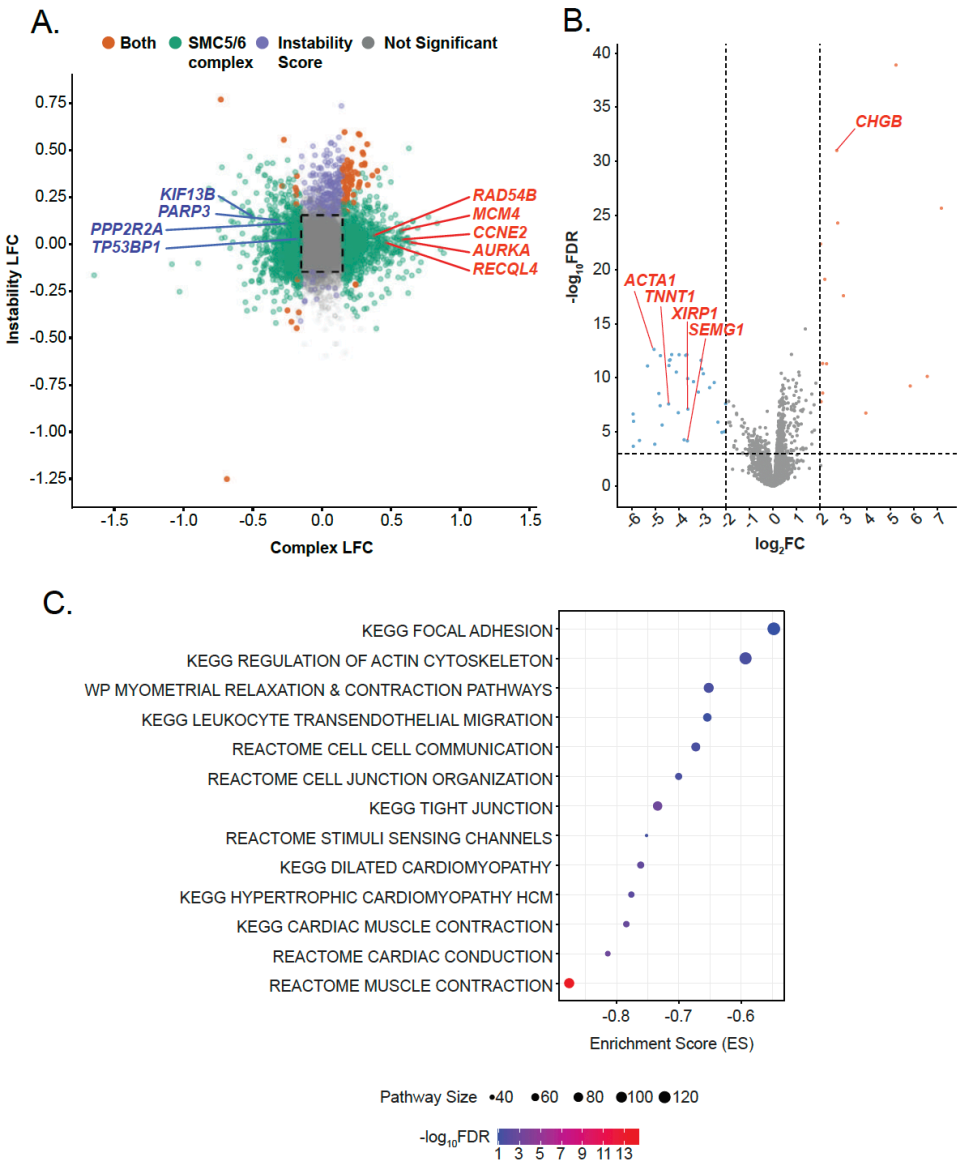


Supplementary Figure 2.S 3 (A) Stratification of NSMCE2 status within the MYC amplified patients in breast cancer. Kaplan-Meier (KM) plot shows the comparison between NSMCE2 wild-type and amplified patients. All patients in this analysis have MYC amplification. NSMCE2 amplified patients have poor prognosis (log-rank p-value=0.035) within the MYC amplified group, indicating

that NSMCE2 has an independent effect on patient survival. (B) Kaplan-Meier (KM) plot for SMC5/6 complex states in cross-cohort breast cancer dataset. Patients were stratified into SMC5/6 wild-type, amplification and mutation groups. Amplification and mutation both have statistically significant negative impact on patient survival (log-rank p-value=8.7x10-9). (C-F) SMC5/6 complex amplification is associated with genome stability and reduced survival in breast cancer patients. Distribution of (C) aneuploidy and (D) ploidy score in breast cancer patients. Data are stratified by SMC5/6 complex and NSMCE2 amplification status (patient with non-amplification genomic alteration where removed). Samples carrying SMC5/6 complex amplification have higher aneuploidy and ploidy score (two-sided Wilcoxon rank-sum test $p < 0.05$). Kaplan-Meier (KM) plots of overall survival in breast cancer, stratified by (E) SMC5/6 complex and (F) NSMCE2 amplification status. Samples with amplification are shown in red and wild type in gray. The bar charts at the bottom of the KM plots represent the number of patients at risk. Amplification in SMC5/6 complex leads to poor overall survival (log-rank test p-value < 0.05).

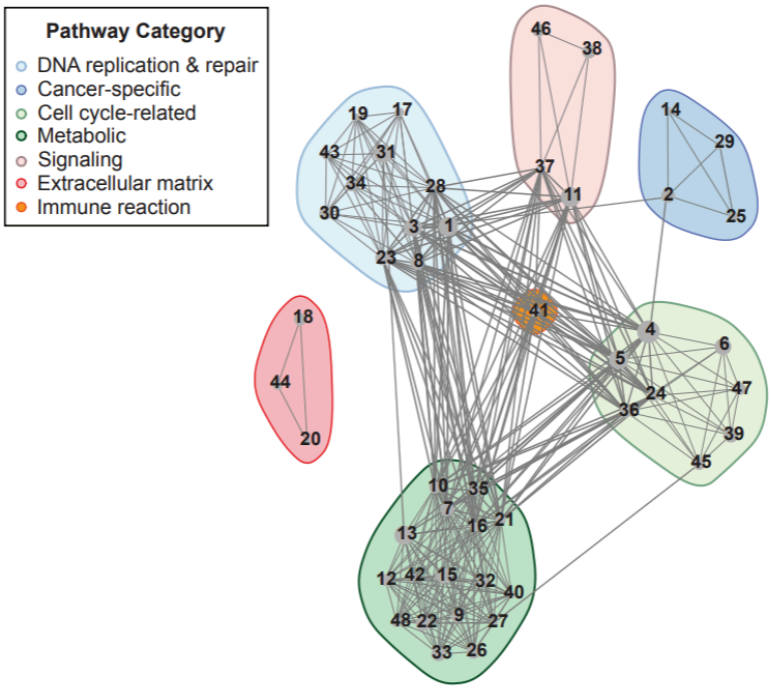


Supplementary Figure 2.S 4: Survival analysis of TCGA ovarian cancer data. The Kaplan-Meier (KM) plot of the progression free survival in the TCGA ovarian cohort stratified by genomic alteration in (A) SMC5/6 complex and (B) NSMCE2 gene. (C) and (D) shows KM plot of the progression free survival for subset of ovarian cancer patients (harboring TP53 mutation, age >55 and have undergone radiotherapy).



Supplementary Figure 2.S 5: (A) Transcriptional changes linked to alterations in the SMC5/6 complex are independent of genomic architecture aberrations. In the plot, each dot represents a gene, the log fold change (LFC) value in the SMC5/6 complex is shown in the x-axis and the y-axis shows LFC from genomic instability. Only a small number of the genes have significant LFC in both SMC5/6 complex and genomic instability groups (shown in red color). Majority of the genes are significant only in SMC5/6 complex alteration (shown in green color). (B) Volcano plot of differentially expressed genes between samples with the SMC5/6 complex alteration and wild type. Genes with absolute log2FC > 2 and FDR <0.001 were considered significant. Significantly downregulated genes are indicated by blue color and significantly upregulated genes are indicated by orange color. (C) Dot plot of gene-set enrichment analysis (GSEA). All pathways with significant enrichment (FDR < 0.15) are shown.

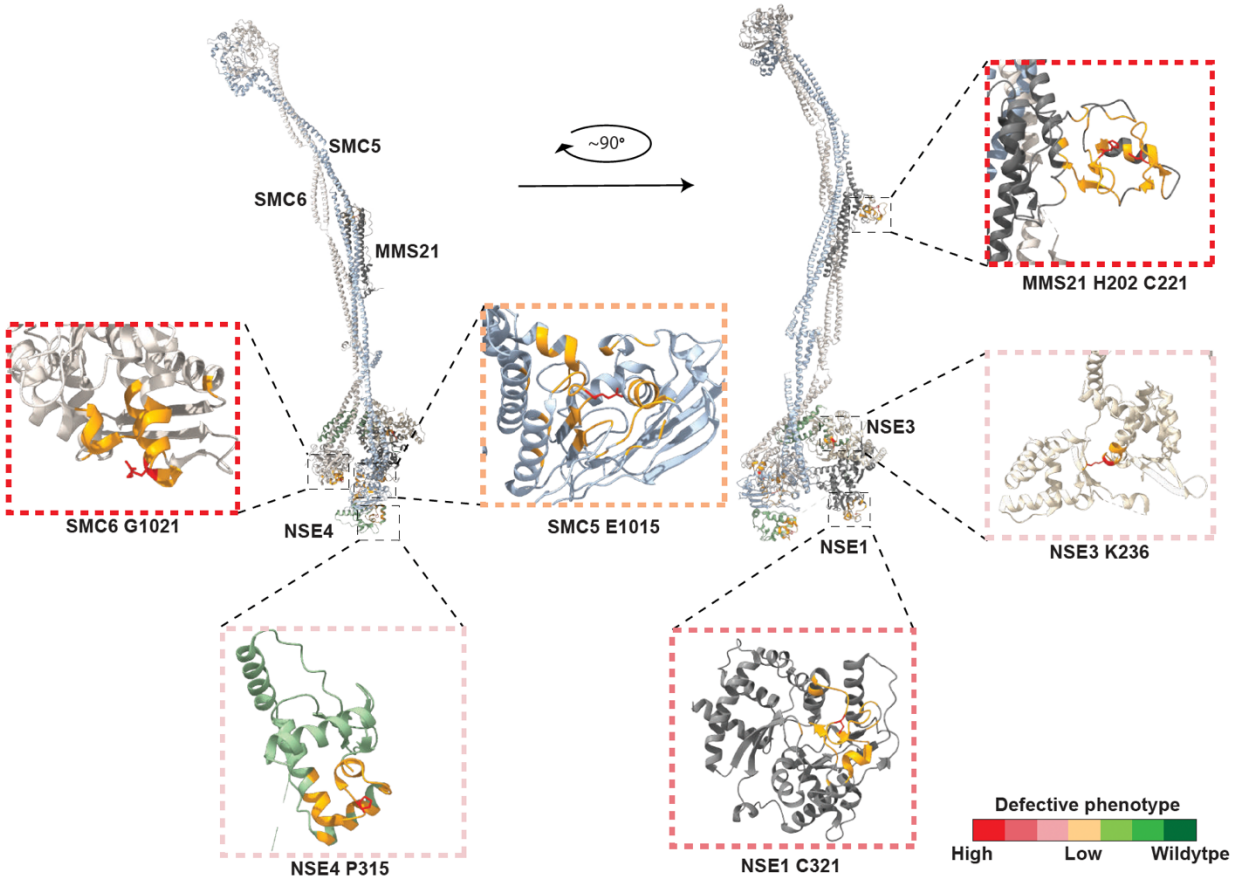
A.



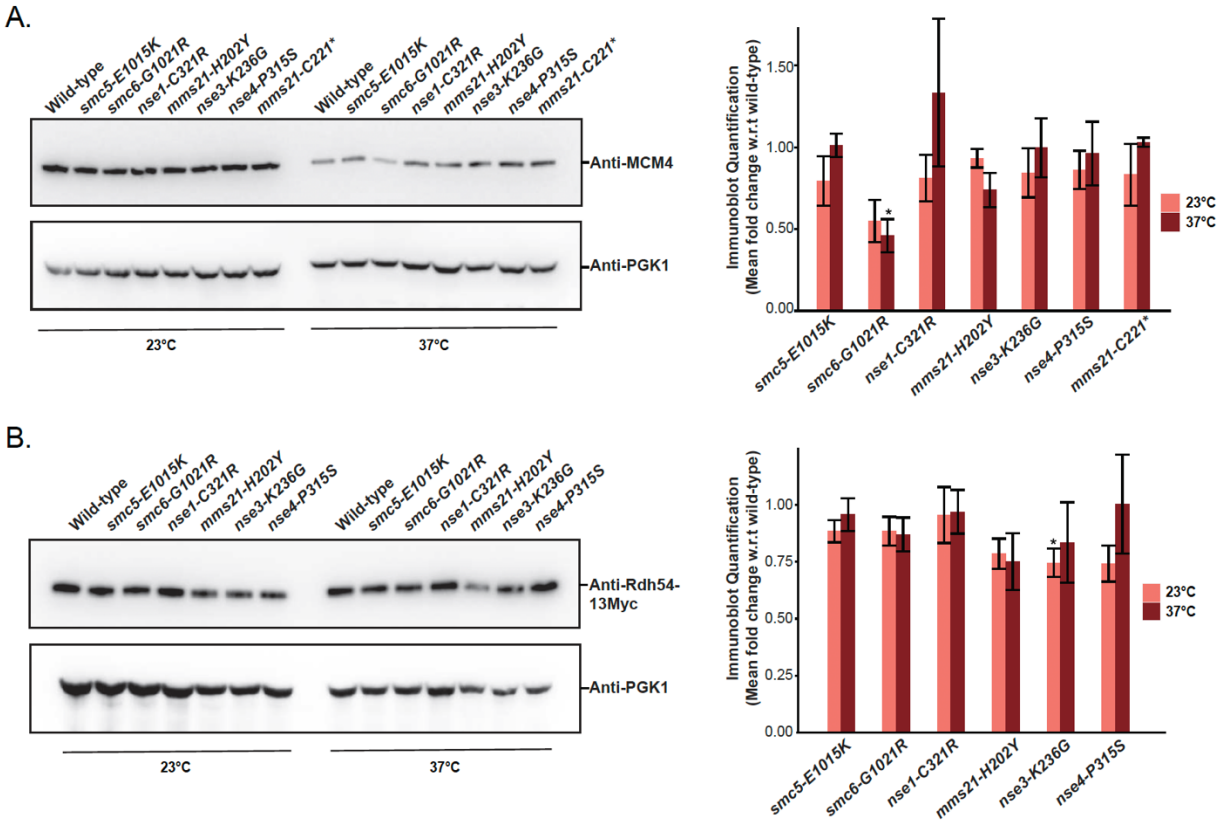
B.

	Pathway Name	Pathway Type	Enrichment Score	-log$_{10}$FDR
1	REACTOME DNA REPLICATION	DNA replication & repair	0.7537467	9.567670
2	WP RETINOBLASTOMA GENE IN CANCER	Cancer-specific	0.7890882	8.855744
3	REACTOME SYNTHESIS OF DNA	DNA replication & repair	0.7496165	8.828013
4	REACTOME MITOTIC G1 PHASE AND G1 S TRANSITION	Cell Cycle-related	0.6873735	7.677484
5	REACTOME G2 M CHECKPOINTS	Cell Cycle-related	0.7122282	7.291808
6	REACTOME MITOTIC SPINDLE CHECKPOINT	Cell Cycle-related	0.7277407	7.178761
7	REACTOME APC C MEDIATED DEGRADATION OF CELL CYCLE PROTEINS	Metabolic	0.7370875	6.280264
8	REACTOME DNA REPLICATION PRE INITIATION	DNA replication & repair	0.7417896	6.001359
9	KEGG DRUG METABOLISM CYTOCHROME P450	Metabolic	-0.7407569	5.349721
10	REACTOME APC C CDH1 MEDIATED DEGRADATION OF CDC20	Metabolic	0.7230583	4.720247
11	REACTOME HOST INTERACTIONS OF HIV FACTORS	Signaling	0.6427130	4.720247
12	KEGG METABOLISM OF XENOBIOTICS BY CYTOCHROME P450	Metabolic	-0.7069010	4.342457
13	REACTOME DNA DOUBLE STRAND BREAK REPAIR	Metabolic	0.6364392	4.342457
14	WP METABOLIC REPROGRAMMING IN COLON CANCER	Cancer-specific	0.7989275	4.194672
15	REACTOME MITOCHONDRIAL TRANSLATION	Metabolic	0.6650564	3.908037
16	REACTOME METABOLISM OF POLYAMINES	Metabolic	0.7395497	3.594516
17	REACTOME TRNA PROCESSING IN THE NUCLEUS	DNA replication & repair	0.7400359	3.528095
18	REACTOME ECM PROTEOGLYCANS	Extracellular Matrix	-0.6800695	3.470672
19	REACTOME SUMOYLATION OF DNA REPLICATION PROTEINS	DNA replication & repair	0.7813509	3.470672
20	NABA BASEMENT MEMBRANES	Extracellular Matrix	0.7763723	3.392350
21	REACTOME SCF SKP2 MEDIATED DEGRADATION OF P27 P21	Metabolic	0.7097036	3.274793
22	REACTOME REGULATION OF TP53 ACTIVITY THROUGH PHOSPHORYLATION	Metabolic	0.6304529	3.126751
23	REACTOME REGULATION OF MRNA STABILITY BY PROTEINS THAT BIND AU RICH ELEMENTS	DNA replication & repair	0.6298003	3.089443
24	REACTOME G1 S DNA DAMAGE CHECKPOINTS	Cell Cycle-related	0.6771099	2.951101
25	REACTOME TELOMERE MAINTENANCE	Cancer-specific	0.6923908	2.861873
26	REACTOME PEPTIDE HORMONE METABOLISM	Metabolic	-0.6252232	2.804659
27	REACTOME NUCLEAR ENVELOPE BREAKDOWN	Metabolic	0.7340431	2.801483
28	REACTOME AUF1 HNRNP D0 BINDS AND DESTABILIZES MRNA	DNA replication & repair	0.7103213	2.757939
29	REACTOME EXTENSION OF TELOMERES	Cancer-specific	0.7323002	2.701105
30	REACTOME ANTIVIRAL MECHANISM BY IFN STIMULATED GENES	DNA replication & repair	0.6297248	2.666523
31	WP DNA REPAIR PATHWAYS FULL NETWORK	DNA replication & repair	0.5847650	2.660626
32	REACTOME DISEASES ASSOCIATED WITH O GLYCOSYLATION OF PROTEINS	Metabolic	-0.6541209	2.573242
33	REACTOME GLUCOSE METABOLISM	Metabolic	0.6000504	2.552662
34	REACTOME SUMOYLATION OF DNA DAMAGE RESPONSE AND REPAIR PROTEINS	DNA replication & repair	0.6364556	2.489598
35	REACTOME DEGRADATION OF DVL	Metabolic	0.6877446	2.480979
36	REACTOME THE ROLE OF GTSE1 IN G2 M PROGRESSION AFTER G2 CHECKPOINT	Cell Cycle-related	0.6294432	2.378559
37	REACTOME NEGATIVE REGULATION OF NOTCH4 SIGNALING	Signaling	0.6768670	2.323588
38	KEGG ECM RECEPTOR INTERACTION	Signaling	-0.5869993	2.266146
39	REACTOME CELLULAR RESPONSE TO HEAT STRESS	Cell Cycle-related	0.5809482	2.247928
40	WP GLYCOLYSIS AND GLUCONEOGENESIS	Metabolic	0.7189077	2.247928
41	REACTOME DECTIN 1 MEDIATED NONCANONICAL NF KB SIGNALING	Immune Reaction	0.6416723	2.231723
42	REACTOME SNRNP ASSEMBLY	Metabolic	0.6934485	2.218336
43	REACTOME RNA POLYMERASE II TRANSCRIPTION TERMINATION	DNA replication & repair	0.6730166	2.217103
44	REACTOME ASSEMBLY OF COLLAGEN FIBRILS AND OTHER MULTIMERIC STRUCTURES	Extracellular Matrix	-0.6394944	2.196765
45	REACTOME MITOTIC PROPHASE	Cell Cycle-related	0.6175087	2.190837
46	REACTOME REGULATION OF INSULIN SECRETION	Signaling	-0.6027229	2.170486
47	WP DNA IRDAMAGE AND CELLULAR RESPONSE VIA ATR	Cell Cycle-related	0.6071791	2.170486
48	WP OXIDATION BY CYTOCHROME P450	Metabolic	-0.6249351	2.145718

Supplementary Figure 2.S 6: Network visualization of pathways associated with genomic alteration in the SMC5/6 complex (related to Figure 2.5C). (A) Each node represents a pathway and the edges between two nodes indicate the number of genes shared between the pathways. (B) Corresponding pathway name associated with the labeled node and pathway enrichment score and p-values.



Supplementary Figure 2.S 7: Structural model of the SMC5/6 complex summarizing the impact of mutations with the strongest proliferation effects. The severity of the phenotypes is color-coded in the frames surrounding the magnified/detailed view of each mutation (i.e., red representing the most defective phenotype, green representing wild-type behavior). The amino-acid positions of the mutations created in this study are labeled in red along with their closest 5Å region in yellow in the magnified view (inset). The crystal structure of the SMC5/6 complex structure used is PDB 7QCD (https://www.rcsb.org/structure/7QCD).



Supplementary Figure 2.S 8: Effect of temperature on protein abundance of (A) Mcm4 and (B) Rdh54 in yeast strains carrying cancer-related mutations in the members of the Smc5/6 complex. Cell lysates of yeast grown at the indicated temperatures were resolved by SDS-PAGE and processed for immunoblot analysis. The positions of Mcm4 and Rdh54 and loading control-phosphoglycerate kinase 1 (PGK1) bands are shown on the right of the blots. Quantitative analysis of Mcm4 and Rdh54 abundance determined by immunoblot are shown on the right of each blot. The bar graph reports the mean protein abundance of Mcm4 and Rdh54 in yeast strains carrying mutated Smc5/6 complex relative to yeast expressing wild-type Smc5/6 complex ± SEM for 3 independent experiments. Asterisk signifies p-value ≤ 0.05 (multiple t-tests were performed comparing the mutants to the wild-type for either temperature, p-values were corrected by Bonferroni multiple correction method).

Table 2.S 1: Yeast strains used in this study.

Figure	Strain name	Relevant genotype details
Figure 2.7	D4107	*MATa*
	D6837	*MATa smc5-6=3HA:: T_{ADH1}::HIS3MX6*
	D7389	*MATa mms21 C221*:: T_{ADH1}::kanMX6*
	D7639	*MATa smc5-E1015K:: T_{ADH1}::kanMX6*
	D7640	*MATa smc6-G1021R:: T_{ADH1}::kanMX6*
	D7684	*MATa nse1-C321R:: T_{ADH1}::kanMX6*
	D7671	*MATa mms21-H202Y:: T_{ADH1}::URAMX6*
	D7641	*MATa nse3-K236G:: T_{ADH1}::kanMX6*
	D7686	*MATa nse4-P315S:: T_{ADH1}::kanMX6*
Figures 2.8	D3	*MATa / MATα*
	D7397	*MATa / MATα SMC5/ smc5-G557*:: T_{ADH1}::URA3MX6*
	D7399	*MATa / MATα SMC5 / smc5 R973*:: T_{ADH1}::URA3MX6*
	D7401	*MATa / MATα SMC5/ smc5 E967*:: T_{ADH1}::URA*
	D7378	*MATa / MATα SMC6/ smc6 E903*:: T_{ADH1}::kanMX6*
	D7540	*MATa / MATα NSE1/ nse1 H316*:: T_{ADH1}::kanMX6*
	D7370	*MATa / MATα MMS21/ mms21 C221*:: T_{ADH1}::KanMX6*
	D7372	*MATa / MATα NSE3/ nse3 K236*:: T_{ADH1}::kanMX6*
	D7384	*MATa / MATα NSE3/ nse3 Y239*:: T_{ADH1}::kanMX6*
	D7380	*MATa / MATα NSE3/ nse3 H290*:: T_{ADH1}::kanMX6*
	D7374	*MATa / MATα NSE4/ nse4 E265*:: T_{ADH1}::kanMX6*
	D7376	*MATa / MATα NSE4/ nse4 K299*:: T_{ADH1}::kanMX6*
	D6183	*MATa / MATα SMC5/ smc5Δ:: T_{ADH1}::HIS3MX6*
	D8108	*MATa / MATα SMC5/ smc5-6=3HA:: T_{ADH1}::HIS3MX6*

	D7514	*MATa / MATα SMC5/ smc5-E1015K:: T_{ADH1}::kanMX6*
	D7429	*MATa / MATα SMC6/ smc6-G1021R:: T_{ADH1}::kanMX6*
	D7627	*MATa / MATα NSE1/ nse1-C321R:: T_{ADH1}::::kanMX6*
	D7487	*MATa / MATα MMS21/ mms21-H202Y:: T_{ADH1}::URAMX6*
	D7450	*MATa / MATα NSE3/ nse3-K236G:: T_{ADH1}::kanMX6*
	D7631	*MATa / MATα NSE4/ nse4-P315S:: T_{ADH1}::kanMX6*

Asterisk (*) signifies the introduction of a stop codon at the position encoding the specified amino-acid residue.

Table 2.S 2: Mutations tested in yeast and their corresponding human mutations.

Human protein	Truncation mutation in human	Truncation mutation in yeast	Type/Location
SMC5	*SMC5-G561**	*SMC5/smc5-G557**	Conserved/Flexible-hinge
	*SMC5-R978**	*SMC5 /smc5-R973**	Conserved/Frequent/ATP-head
	*SMC5-R972**	*SMC5/smc5-E967**	Frequent/ATP-head
SMC6	*SMC6-E870**	*SMC6/smc6-E903**	Conserved/Frequent/ATP-head
NSMCE1	*NSMCE1-S222**	*NSE1/nse1-H316**	Frequent/Zinc finger
NSMCE2	*NSMCE2-C210**	*MMS21/mms21-C221**	Conserved/SP-Ring-domain
NSMCE3	*NSMCE3-R229**	*NSE3/nse3-K236**	Frequent/WH-B-domain
	*NSMCE3-Y232**	*NSE3/nse3-Y239**	Conserved/WH-B-domain
	*NSMCE3-Y282**	*NSE3/nse3-H290**	Frequent/ Outside WH-B-domain
EID3	*EID3-E197**	*NSE4/nse4-E265**	Conserved/outside WH-domain
NSMCE4A	*NSMCE4A-R279**	*NSE4/nse4-K299**	Frequent/WH-domain
Human protein	**Point mutation in human**	**Point mutation in yeast**	**Type/Location**
SMC5	*SMC5-E1020K/Q*	*smc5-E1015K*	Conserved/Frequent/ATP-head
SMC6	*SMC6-G989R*	*smc6-G1021R*	Conserved/ATP-head

NSMCE1	*NSMCE1-C228R*	*nse1-C321R*	Conserved/Zinc-finger
NSMCE2	*NSMCE2-H187Y*	*mms21-H202Y*	Conserved/SP-Ring-domain
NSMCE3	*NSMCE3-R229G*	*nse3-K236G*	Frequent/WH-B-domain
EID3	*EID3-P249S*	*nse4-P315S*	Conserved/Frequent/WH-B-domain

Chapter 3: Manuscript #2

The Smc5/6 complex counteracts R-loop formation at highly transcribed genes in cooperation with RNase H2

Manuscript submitted and in revision at ***eLife***

Shamayita Roy[1], Hemanta Adhikary[1], Sarah Isler [1] & Damien D'Amours[1,2]

[1] Ottawa Institute of Systems Biology, Department of Cellular and Molecular Medicine, University of Ottawa, Roger Guindon Hall, 451 Smyth Rd, Ottawa, ON, K1H 8M5, Canada

[2] Corresponding author: damien.damours@uottawa.ca

3.1 ABSTRACT

The R-loop is a common transcriptional by-product that consists of an RNA-DNA duplex joined to a displaced strand of genomic DNA. While the effects of R-loops on health and disease are well established, there is still an incomplete understanding of the cellular processes responsible for their removal from eukaryotic genomes. Here, we show that a core regulator of chromosome architecture —the Smc5/6 complex— plays a crucial role in the degradation of R-loop structures formed during gene transcription. Consistent with this, mutants defective in the Smc5/6 complex and enzymes involved in R-loop resolution show strong synthetic interactions and accumulate high levels of RNA-DNA hybrid structures in their chromosomes. Importantly, we demonstrate that the Smc5/6 complex recognizes specific types of RNA-DNA hybrid structures *in vivo* and promotes the degradation of R-loops by RNase H enzymes. Collectively, our results reveal a crucial role for the Smc5/6 complex in the removal of toxic R-loops formed at highly transcribed genes and telomeres.

3.2 INTRODUCTION

The maintenance of genome stability is a primordial function that ensures the proper development and homeostasis of all living organisms[1–4]. Successful maintenance of genomic integrity requires constant monitoring and repair of DNA lesions because genomes are under constant attack from endogenous sources of DNA damage[5–7]. One of the most common sources of endogenous DNA damage is the formation of RNA-DNA hybrid structures in chromosomes. During transcription, nascent RNA transcripts can re-anneal to complementary DNA strands producing an RNA-DNA hybrid and displace the template strand, creating an obstacle to the progression of the DNA replication machinery. The resulting RNA-DNA hybrid and displaced ssDNA segment —together termed the R-loop— are highly deleterious for genome integrity[8–11]. RNA-DNA hybrid structures can also be formed under a variety of physiological conditions in eukaryotic genomes and play important roles in cell physiology and regulate genome dynamics. Common physiological roles of RNA-DNA hybrids include immunoglobulin class switching recombination of B cells in vertebrates[12,13], mitochondrial DNA replication[14,15], bacterial plasmid replication[16,17], and CRISPR-Cas9 gene editing[18,19]. Moreover, RNA-DNA hybrids co-ordinate specific regulatory steps in transcription initiation and termination[20–22], telomere homeostasis[23,24], and gene expression[25–29]. However, the formation of unprogrammed or non-physiological RNA-DNA hybrid structures can interfere with DNA replication-related processes, resulting in replicative stress and the formation of DNA double-strand breaks (DSBs)[25,30–34]. Importantly, exposed ssDNA in the R-loop can be cleaved by different endonucleases leading to DNA breaks and/or mutagenic events and can also adopt harmful secondary structures[35,36]. R-loop-induced DNA damage and genomic rearrangements have been linked to various disease

states in humans. Examples include trinucleotide repeat-associated diseases, auto-immune disorders, neurological disorders, and cancer, although it is not currently known whether R-loops play a causative or consequential role in such diseases[26,37].

To mitigate the toxic consequences associated with the presence of unprogrammed RNA-DNA hybrids in chromosomes, several cellular mechanisms work in concert to prevent their formation, and when they do accumulate, remove them from eukaryotic genomes. For instance, the Ribonuclease H (RNase H) family of enzymes plays a central role in degrading the RNA moiety of R-loops created under a variety of genomic conditions[38,39] (Figure 3.1A). Recently, it was shown that the RNase DICER can also cleave the RNA moiety of the R-loops in higher eukaryotes[40]. In addition, factors associated with transcription and mRNA biogenesis, RNA-DNA helicases, topoisomerases, chromatin remodelers, and several DNA repair enzymes are known to be involved alongside RNase H in preventing the accumulation of R-loop structures in eukaryotic chromosomes[26]. However, a definitive understanding of the sequence of events and exact molecular mechanisms responsible for the repair of R-loops/RNA-DNA hybrids remains to be established.

Work from our laboratory and other researchers provides hints of a possible involvement of structural maintenance of chromosomes (SMC)-type complexes in RNA-DNA hybrid metabolism[41–44]. SMC complexes are effectors of large-scale changes in chromosome organization and include three evolutionarily conserved enzyme complexes: condensin, cohesin, and the Smc5/6 complex[45]. The Smc5/6 complex is a particularly intriguing member of this family

because its function —unlike that of cohesin and condensin— is primarily concerned with DNA repair, and yet its exact contribution to this process is not fully understood. Recent evidence suggests that the Smc5/6 complex is a DNA compacting enzyme that acts *in vivo* by regulating local chromatin domains containing unusual DNA structures that can lead to replication stress and/or DNA damage[42,46–48]. Interestingly, genome-wide screens as well as targeted genetic analyses revealed synthetic interactions among a subset of mutants affecting Smc5/6 complex components and RNA-DNA hybrid detoxification enzymes[41,49–52]. Consistent with this, we have previously shown that the Smc5/6 complex can bind to short RNA-DNA duplexes with high affinity and specificity[42]. Together, these results raise the intriguing possibility that the Smc5/6 complex might be involved in the detection and/or processing of toxic R-loops formed in eukaryotic genomes.

Here, we show that the Smc5/6 complex is directly involved in the repair of unscheduled R-loops generated by a diverse set of genomic transactions. In particular, we demonstrate that the Smc5/6 complex binds strongly to R-loop structures formed during active gene transcription and promotes RNase H2-mediated degradation of the RNA component of R-loops. Our results unravel a hitherto unanticipated role for the Smc5/6 complex in the removal of RNA structures from chromosomes, an essential function for the maintenance of genome integrity and cell fitness.

3.3 RESULTS

The Smc5/6 complex collaborates with RNase H enzymes in the maintenance of genome integrity

In eukaryotes, two partially overlapping enzymes mediate the degradation of R-loops; RNase H1 and RNase H2 (Figure 3.1A)[38,39]. To determine whether the Smc5/6 complex contributes to R-loop repair in proliferating cells, we introduced the *smc6-9* and *nse4-4* alleles of the Smc5/6 complex (Figure 3.1B) into budding yeast strains defective for RNase H1 (*rnh1Δ*) and RNase H2 (*rnh201Δ/rnh202Δ/rnh203Δ*) activity. We used the *smc6-9* and *nse4-4* alleles for our analysis because they correspond to moderate and a strong temperature-sensitive mutants of the Smc5/6 complex, respectively, and inactivate its DNA repair activity in a general/non-specific manner[53,54]. For simplicity, we will refer to strains defective in both Smc5/6 complex and RNase H activities as double mutants and their parent strains as single mutants (i.e., even if the corresponding strains carry more than one mutant allele).

The proliferation rate of the Smc5/6-RNase H double mutants was compared to that of parental single mutants at various temperatures (23°C, 30°C, and 32°C) and in the presence of DNA-damaging agents (hydroxyurea [HU], 4-nitroquinoline 1-oxide [4NQO], and methyl methanesulfonate [MMS]) (Figure 3.1D and 3.S1). Loss of both Smc5/6 and RNase H activities resulted in impaired proliferation even under optimal growth conditions, as illustrated by the growth patterns of haploid spores after sporulation and dissection of heterozygous diploid strains (Figure 3.1C). Competitive proliferation assays revealed that double mutant strains exhibit heightened temperature sensitivity at 30°C and 32°C compared to their corresponding parental

strains (Figure 3.1D). Additionally, most of the double mutants were more sensitive than single mutant strains when exposed to DNA-damaging conditions (e.g., see HU at 23°C in Figure 3.1D; MMS and 4NQO in Figure 3.S1A). Overall, the proliferation defect associated with Smc5/6 complex mutations was substantially exacerbated by the loss of RNase H activity (Figure 3.1D and 3.S1A). A synthetic growth defect was observed when Smc5/6 complex mutations were combined with RNaseH2 enzyme inactivation (Figure 3.S1B), suggesting that the Smc5/6 complex collaborates specifically with RNase H2 in the maintenance of genome integrity.

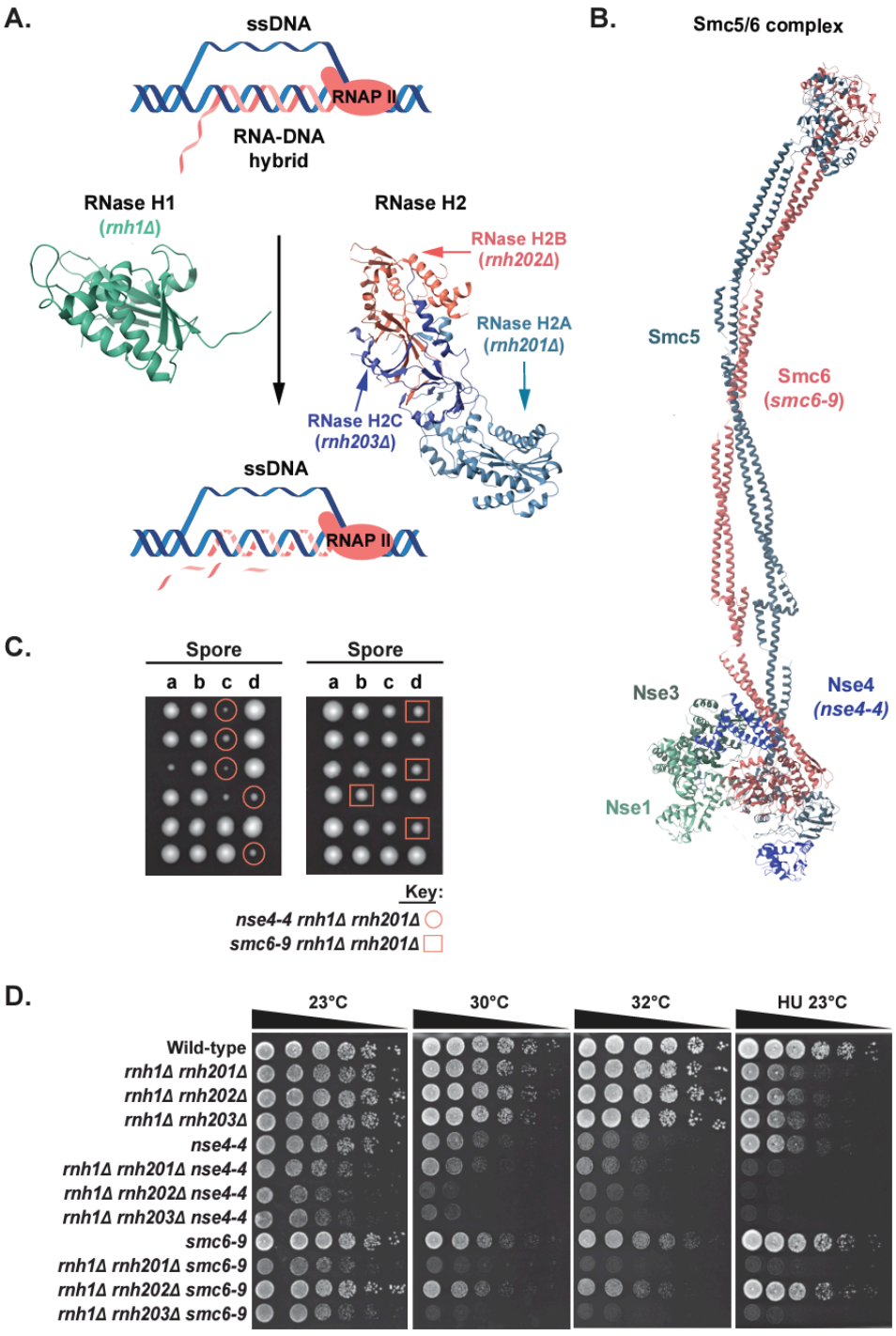


Figure 3. 1: Synthetic enhancement of RNase H enzyme defects by mutations affecting Smc5/6 complex activity. (A) Schematic model showing the mechanism of RNase H mediated RNA-DNA hybrid degradation. The crystal structures of the RNase H1 and RNase H2 complex used in this representation are PDB: 2QK9[55] and PDB: 3PUF[56], respectively. (B) Schematic representation of the Smc5/6 complex showing its subunits and the corresponding mutant alleles used in this study. The crystal structure of the Smc5/6 complex used in this representation is PDB: 7QCD[57]. (C)

Growth of *nse4-4 rnh1Δ rnh201Δ* and *smc6-9 rnh1Δ rnh201Δ* haploid spores after sporulation and germination of heterozygous diploid strains. The viability of the haploid spores was scored after 3 days of germination. (D) The proliferation capacity of combination mutants affecting Smc5/6 complex and RNase H activity was monitored after dilution on solid medium and growth under various conditions (indicated on top of the growth medium). Concentration of HU used was 12.5mM. YPD 23°C, 30°C, 32°C, and HU plates were grown in temperature-controlled incubators for ~48 hrs, ~28hrs, ~26hrs, and ~72 hours respectively, before scanning the plates.

Simultaneous loss of Smc5/6 complex and RNase H activity exacerbates RNA-DNA hybrid accumulation in chromosomes

Next, we wanted to evaluate whether the growth defect observed in Smc5/6-RNase H double mutants is due to defective RNA-DNA hybrid removal. We quantified RNA-DNA hybrid foci by indirect immunofluorescence on chromosome spreads stained with the S9.6 antibody[58]. Remarkably, we detected a large increase in the numbers of S9.6 foci on chromatin spreads prepared from double mutants compared to those from single mutant strains (Figures 3.2 and 3.S2A). Whereas RNase H mutants alone typically show a few RNA-DNA hybrid/S9.6 foci per nucleus, inactivating Smc5/6 components in this genetic background resulted in the accumulation of more than 10 foci per nucleus. The presence of S9.6 foci was highest in the *nse4-4 rnh1Δ rnh201Δ* and *smc6-9 rnh1Δ rnh201Δ* strains, consistent with the fact that *rnh201Δ* represents the deletion of the catalytic subunit of RNase H2 (Figure 3.2A). *nse4-4* and *smc6-9* mutations alone did not result in a statistically significant increase in RNA-DNA foci formation, indicating that cells possess excess R-loop processing capacity when RNase H and other alternate R-loop metabolism pathways are active. Importantly, overexpression of RNase H1 largely suppressed the S9.6 signal on chromatin spreads (Figure 3.2B), indicating that the foci described above are reflective of RNA-DNA hybrid formation in double mutant strains. To further confirm

the accumulation of RNA-DNA hybrids in Smc5/6-RNase H double mutants, we performed a S9.6 antibody-mediated immunoprecipitation assay followed by qPCR analysis, as described in[59]. We focused this analysis on loci that are known to be highly enriched in R-loop formation (Figure 3.3A). Consistent with the chromosome spread analysis, we observed increased RNA-DNA hybrid immunoprecipitation specifically at telomeres and ribosomal genes in the Smc5/6-RNase H double mutant strains compared to the parental strains deficient in only RNase H activity (Figure 3.3A).

In addition to the results presented above that are based on the binding between RNA-DNA hybrids and the S9.6 antibody; we wanted to test if the accumulation of RNA-DNA hybrid structures in the Smc5/6-RNase H double mutant strain could be detected using an alternative approach. Previous studies have established that the single-stranded DNA region of R-loop structures can be directly targeted by various mutagenic enzymes, one of them being the activation-induced cytosine deaminase (AID). This enzyme is highly active on single-stranded DNA during active transcription and creates mutations in DNA by deamination of cytosine and converting cytosine into uracil. This event leads to increased Rad52 foci formation in yeast, a homologous recombination-related phenotype that can be exploited to quantify R-loop abundance upon overexpression of the AID enzyme[60]. In line with our previous results, we observed a substantially higher level of AID-induced Rad52-GFP foci formation in the *nse4-4 rnh1Δ rnh201Δ* mutant strain compared to a wild-type control or the parental single mutants (*rnh1Δ rnh201Δ*; Figure 3.3B, wild-type and *nse4-4*; Figure 3.S2B). These results suggest that the severe growth defects of yeast defective in both RNase H and the SMC5/6 complex can be linked

to an accumulation of R-loops in the chromosomes of these cells.

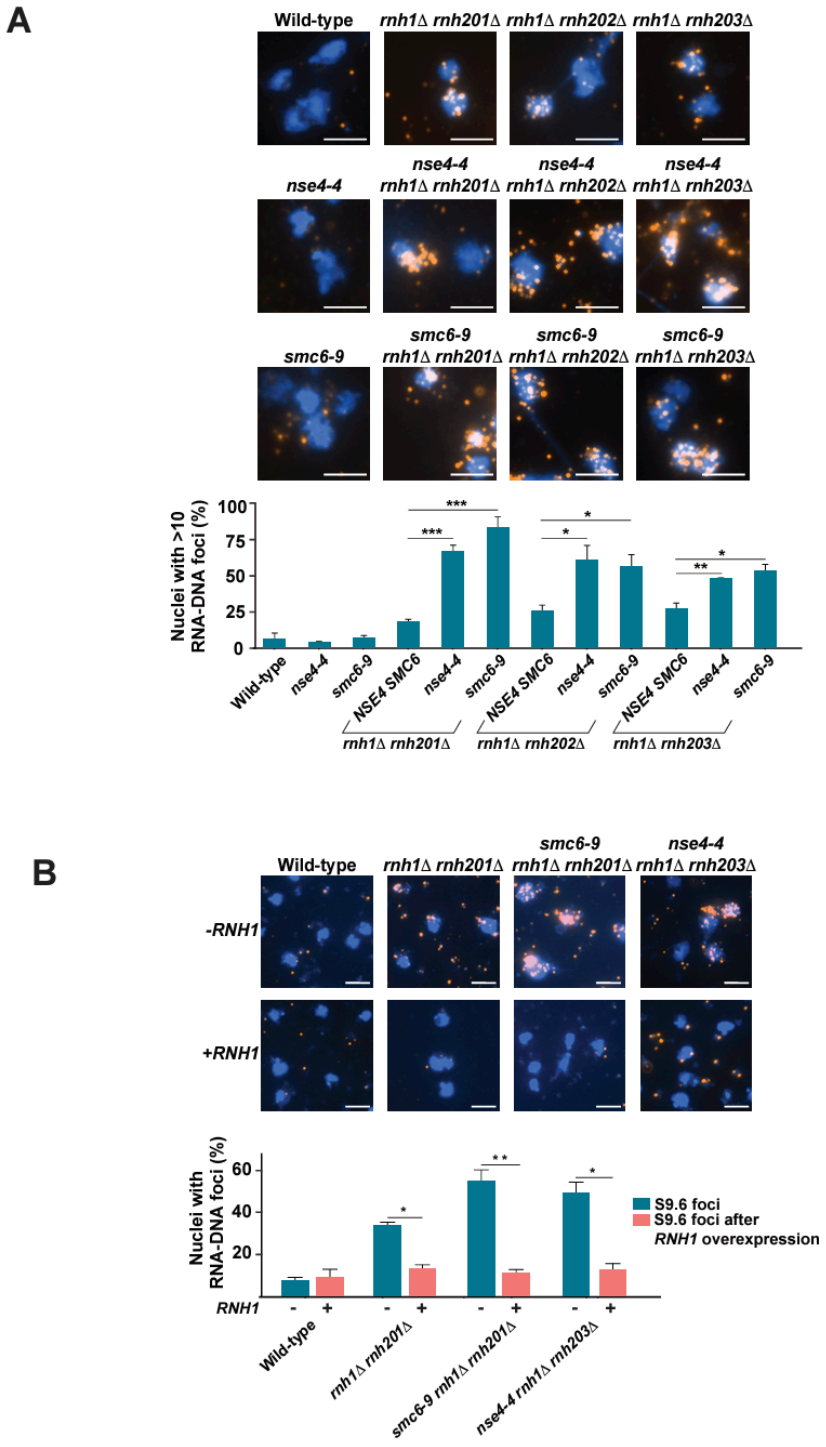


Figure 3. 2: RNA-DNA hybrid accumulation in cells defective for Smc5/6 complex and RNase H activity. (A) The abundance of RNA-DNA hybrids in chromosomes was monitored using the S9.6 antibody by indirect immunofluorescence microscopy on chromosome spreads prepared from

WT, single- and double-mutant yeast strains. Representative spreads are shown, with DNA stained in blue (DAPI) and orange foci representing RNA-DNA hybrid structures detected by the S9.6 antibody. Quantification of nuclei containing S9.6 foci (> 10 foci per nucleus) is shown below the images. 100-200 nuclei were visualized and manually counted for each replicate to obtain the fraction of nuclei with detectable RNA-DNA hybrids. Data represent mean and SE of three independent experiments. *P* (*) < 0.05; *P* (**) < 0.01; *P* (***) < 0.001 (two-tailed Student's t-test). Scale bar, 5 µm. (B) RNA-DNA hybrid were monitored and quantified as described above in the presence and absence of ectopic overexpression of RNase H1 in strains carrying *rnh1Δ rnh201Δ*, *nse4-4 rnh1Δ rnh201Δ* and *nse4-4 rnh1Δ rnh203Δ* mutations. Scale bar, 5 µm.

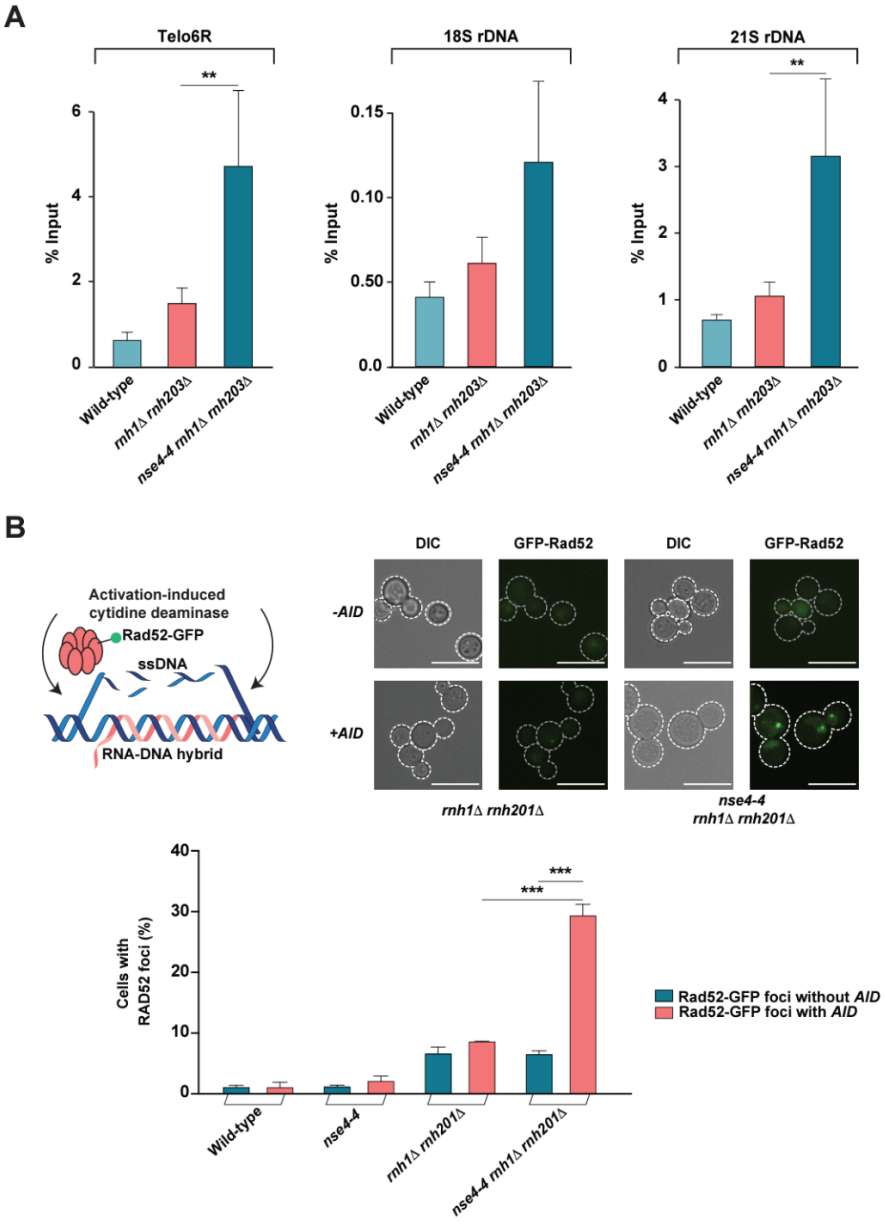


Figure 3. 3: Quantification of R-loop abundance in cells defective for Smc5/6 and RNase H activity. (A) Mutations in the Smc5/6 complex components increase the levels of RNA-DNA hybrids at rDNA genes and telomeres in the absence RNase H activity. DRIP was performed at *TEL06R*, rDNA 18S, and rDNA 21S loci with the S9.6 antibody using genomic DNA prepared from asynchronous cultures of WT, single- and double-mutant yeast strains. Data represents mean and SE of at least 3 independent experiments. *P* (*) < 0.05; *P* (**) < 0.01; *P* (***) < 0.001 (two-tailed Student's t-test). (B) Detection of R-loops using AID-induced Rad52 foci formation. (Top Left) Schematic illustration of AID-induced R-loop mutagenesis and subsequent Rad52 activation. (Top Right) Representative images of cells carrying Rad52-GFP foci in the absence (–AID) or presence of AID (+AID). (Bottom) Quantification of cells showing Rad52-GFP foci after AID overexpression in WT, single- and double-mutant yeast strains. About 100 cells were visualized and manually counted for each replicate to obtain the fraction of cells with detectable Rad52-GFP foci. Data represent mean and SE of three independent experiments. *P* (*) < 0.05; *P* (**) < 0.01; *P* (***) < 0.001 (two-tailed Student's t-test). Scale bar, 10 µm.

The Smc5/6 complex acts on R-loops formed at highly transcribed genes and telomeres

R-loops are most frequently observed at highly transcribed genes but can also be formed at telomeres and near DNA replication forks (Figure 3.4A). To identify the source of R-loops that are substrates for the Smc5/6 complex, we introduced mutant alleles of the Smc5/6 complex in yeast backgrounds that accumulate R-loops (or RNA-DNA hybrid structures) at specific genomic locations.

We first asked if unscheduled R-loops formed during active transcription are natural substrates/targets of the Smc5/6 complex. To test this notion, we used yeast strains defective for the Sen1 helicase (*sen1-1*) and THO complex (*hpr1Δ*), two conditions that lead to very high levels of RNA-DNA hybrids in actively transcribed genes (Figure 3.4A)[61,62]. Synthetic/aggravating interactions with these two mutant conditions is frequently used as a genetic assay to test the contribution of putative effectors of R-loop metabolism (*e.g.*,[63]). Interestingly, deletion of *HPR1*

was synthetic lethal when combined with the *nse4-4* mutation, as evidenced by the growth pattern of haploid spores following sporulation and dissection of a heterozygous diploid strain (Figure 3.4B). Although *hpr1Δ smc6-9* double mutants were viable, the growth defect associated with the *hpr1Δ* mutation was strongly exacerbated in the presence of *smc6-9* at both permissive and restrictive temperatures (Figure 3.4B). *smc6-9* and *nse4-4* alleles also experienced synthetic growth defects when combined with *sen1-1* (see 30°C /32°C and HU/MMS conditions in Figure 3.4B). The *nse4-4 sen1-1* mutant showed defective proliferation even under normal/unchallenged growth conditions, consistent with the more severe temperature sensitivity of this allele compared to that of *smc6-9* (Figure 3.4B). Moreover, we detected a significant increase in the numbers of S9.6 foci on chromatin spreads prepared from *nse4-4 sen1-1* and *smc6-9 hpr1Δ* mutants in comparison to those from the corresponding single mutant strains (Figure 3.S3).Taken together, these genetic interactions suggest that R-loop formed at highly transcribed genes are physiological substrates for the Smc5/6 complex.

Next, we combined *smc6-9* and *nse4-4* mutations with alleles of DNA polymerase ε and Sen1 helicase that increase the formation of RNA-DNA hybrids during DNA replication. Specifically, the *pol2-M644G* mutant exhibits a 10-fold increased ribonucleotide incorporation during DNA replication[64] whereas the *sen1-3* mutation impairs the interaction of Sen1 with the replication machinery, thereby increasing R-loop formation in the vicinity of replication forks (Figure 3.4A)[63]. We did not observe synthetic growth defects in double mutants of these genes with *smc6-9* and *nse4-4* mutations (Figure 3.4B), suggesting that DNA replication-associated RNA-DNA hybrid structures are not substrates for the Smc5/6 complex *in vivo*.

Finally, we investigated whether the Smc5/6 complex interacts with a natural R-loop formed at telomeres; the telomeric repeat-containing RNA [TERRA]-DNA hybrid. To this end, we combined *smc6-9* and *nse4-4* mutations with the *rat1-1* allele defective in the 5' to 3' exonuclease activity responsible for TERRA removal (Figure 3.4A)[24]. The resulting double mutant strains exhibited stronger growth defects than the corresponding single mutants (Figure 3.4B), indicating a role for the Smc5/6 complex in R-loop metabolism at telomeres. This result is consistent with the significant accumulation of increased RNA-DNA hybrids specifically at telomeres in Smc5/6-RNase H double mutants (Figure 3.3A). Combining *rat1-1* and *smc6-9* mutations to *sen1-1* phenocopied the *sen1-1 smc6-9* double mutant (Figure 3.4C), consistent with an involvement of Sen1 in telomeric R-loops metabolism. Overall, these genetic analyses suggest that R-loops formed at highly transcribed genes and telomeres are likely physiological targets for the Smc5/6 complex.

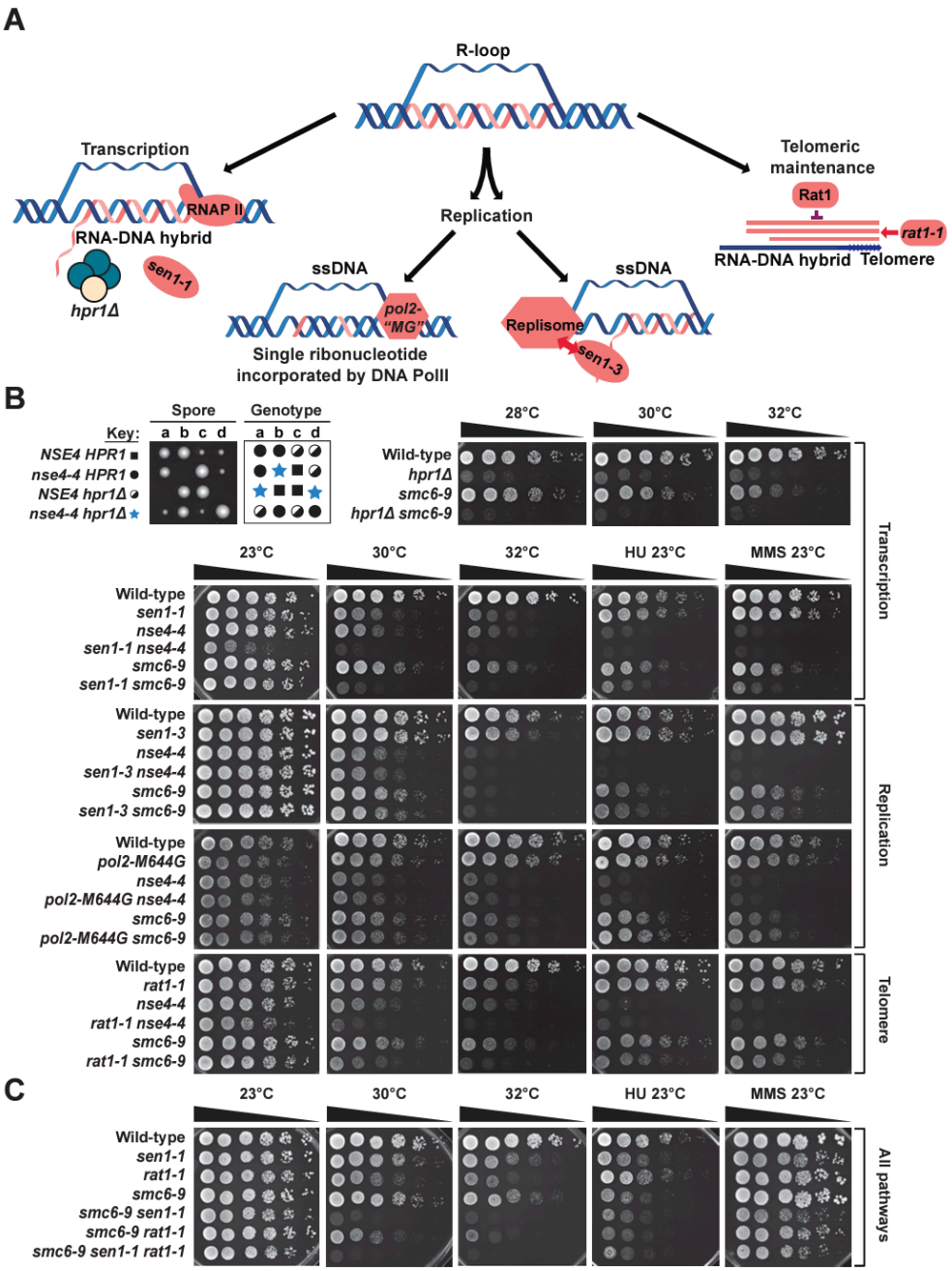


Figure 3. 4: R-loops formed at highly transcribed genes and telomeres are endogenous targets for the Smc5/6 complex. (A) Schematic representation of various cellular mechanisms responsible for RNA-DNA hybrid formation in chromosomes and relevant proteins/mutant implicated in each process. (B–C) Proliferation capacity of yeast strains carrying the specified mutations was monitored by dilution assay as described in Figure 3.1. The growth temperatures and the presence of specific DNA-damaging agents (MMS concentration was 0.005%; HU concentration was 25mM in panel B and 100mM in panel C) in the growth medium are indicated on top of the images. YPD 23°C, 30°C, 32°C, MMS, and HU plates were grown in temperature-controlled

incubators for ~48 hrs, ~28hrs, ~26hrs, ~48hrs, and ~72 hours respectively, before scanning the plates.

The Smc5/6 complex is a high-affinity R-loop-binding enzyme

The role we uncovered above for the Smc5/6 complex prompted us to investigate whether this enzyme was capable of recognizing R-loop structures directly. Extensive research over recent years has revealed that yeast and human Smc5/6 complexes are highly conserved and show strong binding affinities towards DNA substrates that mimic single-stranded (ss)–double-stranded (ds) DNA junctions, supercoiled or catenated DNA, and even branched DNA structures[42,46,47]. We have also established that the human SMC5/6 complex can bind short RNA-DNA duplexes *in vitro*[42], suggesting the enzyme is capable of recognizing the more complex R-loop structure. To test this notion, we purified the human SMC5/6 complex and prepared R-loop and D-loop substrates (Figures 3.5A-B and 3.S4A) to conduct binding experiments by electrophoretic mobility shift assays (EMSAs). The size of R-loop and D-loop substrates used in binding experiments are the same, enabling direct comparison of SMC5/6 complex affinity for these substrates (Figure 3.S4A). We observed that the SMC5/6 complex can bind both R-loop and D-loop structures in EMSA experiments, but its affinity for the R-loop structure is significantly greater (Figure 3.5C). For instance, the SMC5/6 complex can bind the R-loop substrate at a concentration as low as 25nM, as evident from the clear protein/DNA band shift formed in gel at that concentration of enzyme (lane 4; Figure 3.5C), whereas we observed little to no binding to D-loops at 25nM SMC5/6 complex (lane 10; Figure 3.5C). We also observed much higher band shift for R-loop compared to D-loop at 50nM concentration of the SMC5/6 complex (lane 5 and 11; Figure 3.5C). Together, our EMSA experiments indicate that the SMC5/6 complex can

associate with R-loops with high affinity and specificity.

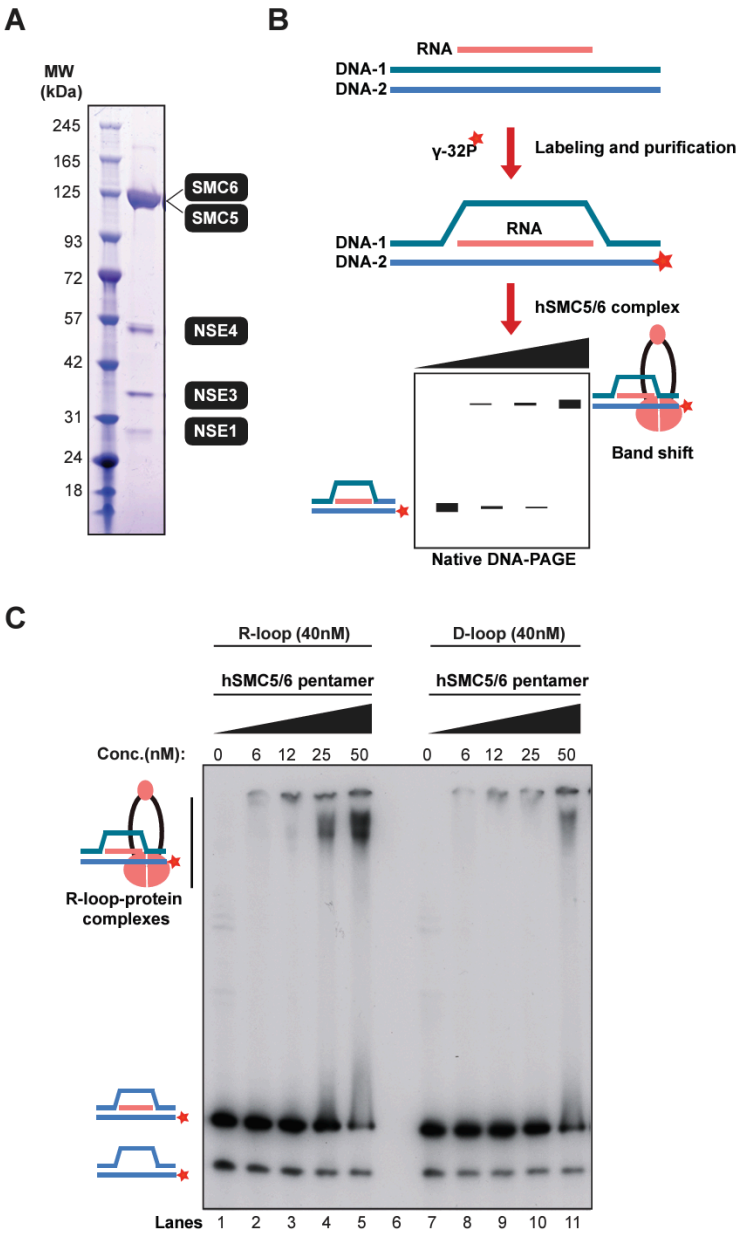


Figure 3. 5: R-loops are high affinity substrates for the Smc5/6 complex. (A) Coomassie blue stained gel showing the purified human SMC5/6 complex used in R/D-loop binding experiments. All the subunits of the SMC5/6 complex migrate in SDS-PAGE at the positions of the native full-length proteins. (B) Schematic representation of the reaction steps for the production of [γ ^{32}P-labeled] R/D-loop substrates and their expected behavior in electromobility shift assays (EMSAs) on a 6% native polyacrylamide gel. Blue and pink strands represent DNA and RNA, respectively, while asterisk indicates the ^{32}P label introduced at the 5'-end of the RNA/DNA strand. (C) EMSA were performed with recombinant human SMC5/6 complex using R-loop and D-loop substrates.

The concentration (nM) of the SMC5/6 complex is indicated above each lane, whereas the positions of unbound substrates and SMC5/6-bound R/D-loop substrates are marked by representative graphics on the side of the gels.

The Smc5/6 complex stimulates the degradation of R-loops by RNase H2

Next, we investigated whether binding of the SMC5/6 complex to an R-loop substrate can affect its degradation and/or stability *in vitro*. RNase H1 and RNase H2 are the primary enzymes responsible for the removal of R-loops in chromosomes, but recent studies suggest that RNase H2 is the only one that acts throughout the cell cycle[39,65]. Consistent with this, we also observed a remarkably strong genetic interaction when combining Smc5/6 complex mutations with RNaseH2 enzyme inactivation (Figure 3.S1B). We therefore tested the impact of the human SMC5/6 complex on the catalytic activity of purified human RNase H2 enzyme in a reconstituted R-loop degradation assay (Figure 3.6A). To this end, we introduced a P^{32} radiolabel on the RNA moiety of our R-loop substrate to allow direct visualization of RNA degradation (Figures 3.6B and 3.S4). We first confirmed that human RNase H2 was able to cleave the radiolabeled RNA in a time and concentration-dependent manner (Figure 3.S4B-C). We also observed that the purified RNase H2 enzyme was not active on a D-loop structure, thereby demonstrating the specificity of the enzyme in our reaction conditions (Figure 3.S4D). Next, we incubated increasing concentrations of the SMC5/6 complex with the R-loop substrate in the absence of RNase H2. Under these conditions, the complex did not induce degradation of the RNA moiety or otherwise affect the stability of the R-loop structure (lanes 2-6; Figure 3.6C). However, in the presence of low levels of RNase H2 enzyme, the same concentrations of SMC5/6 complex induced a major stimulation of R-loop degradation. This led to a rapid accumulation of radiolabeled product at

the bottom of the gel (lanes 8-12; Figure 3.6C), reflecting the nucleolytic processing of the RNA moiety within the R-loop by RNase H2. The stimulation of RNase H2 activity by the SMC5/6 complex was concentration-dependent and evident even at the lowest concentration of the SMC5/6 complex tested in this experiment (1.25nM, lane 8; Figure 3.6C). To test the possibility that the SMC5/6 complex physically binds to RNaseH2 enzyme directly, we performed an in-vitro pull-down assay and found no detectable physical interaction connecting these proteins (Figure 3.S5). Taken together, these results demonstrate that the SMC5/6 complex can promote R-loop degradation by stimulating substrate degradation by the RNase H2 enzyme without the establishment of a strong physical interaction with RNaseH2.

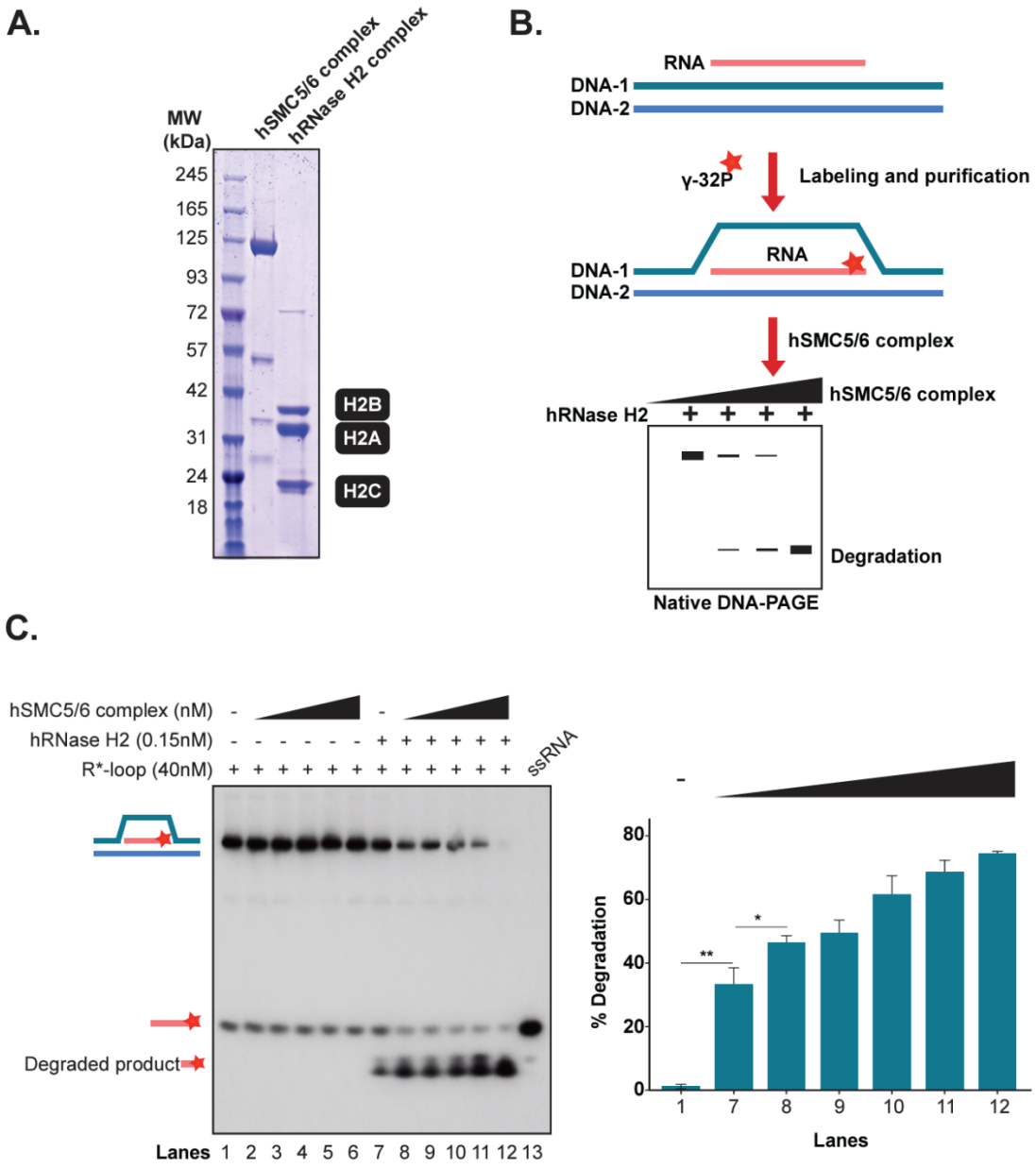


Figure 3. 6: The SMC5/6 complex stimulates the degradation of R-loops by RNase H2. (A) Coomassie blue-stained gel showing the purity of recombinant RNase H2[66] and SMC5/6 complex[42] used in R-loop degradation assays. All the components of the SMC5/6 and RNase H2 holoenzymes migrate in SDS-PAGE at the positions expected for the native/full-length subunits of their respective complexes. (B) Schematic representation of the steps involved in the production of a radiolabeled R*-loop probe and in the RNase H2 degradation assay. Blue and pink strands represent DNA and RNA, respectively, while the asterisk marks the 32P label introduced in the RNA strand of the R-loop structure. (C) R-loop degradation assay conducted in presence of human RNase H2 and increasing concentration of human SMC5/6 complex (1.25nM, 2.5nM, 5nM, 10nM, 20nM). The bar graph (next to the gel) shows the quantification of the degradation assay.

Individual bars report the mean and SE of 4 independent experiments. P (*) < 0.05; P (**) < 0.01; P (***) < 0.001 (Student's t-test).

3.4 DISCUSSION

The incorporation of RNA in chromosomes represents a unique challenge for the stability of eukaryotic genomes because this modification can occur in both normal and pathogenic conditions. Maintaining a finely balanced cycle of R-loop formation and removal in chromosomes is crucial for the overall fitness of cells because altered RNA-DNA hybrid homeostasis can result in DNA damage and genomic instability, ultimately contributing to the development of several pathological conditions[67,68]. Here, we show that the Smc5/6 complex promotes the removal of toxic R-loops in eukaryotic chromosomes. While previous genetic experiments have supported a role for the Smc5/6 complex in the natural regulation of TERRA levels at telomeres[41,69], our study reports the first demonstration that Smc5/6 complex activity is essential for the removal of unscheduled R-loops from the genome. This discovery is significant because non-physiological R-loops represent the most toxic and damaging source of RNA-DNA hybrids for genome stability[25,27,70] and failure to remove these structures exacts a heavy toll on cell fitness. Moreover, we demonstrate for the first time that the Smc5/6 complex can directly recognize R-loops, suggesting an early role in the detection and repair of these structure *in vivo* (see model in Figure 3.7). Consistent with this suggestion, the Smc5/6 complex has been shown by chromatin immunoprecipitation to accumulate at sites that are common R-loop enrichment zones on chromosomes, including the rDNA locus, telomeres, and highly transcribed/difficult to replicate chromosomal loci[71–74]. We showed that inactivation of Smc5/6 components lead to an increase in R-loop formation at several of these loci in the absence of RNase H enzyme activity, a result

that aligns nicely with Smc5/6 complex localization in live cells.

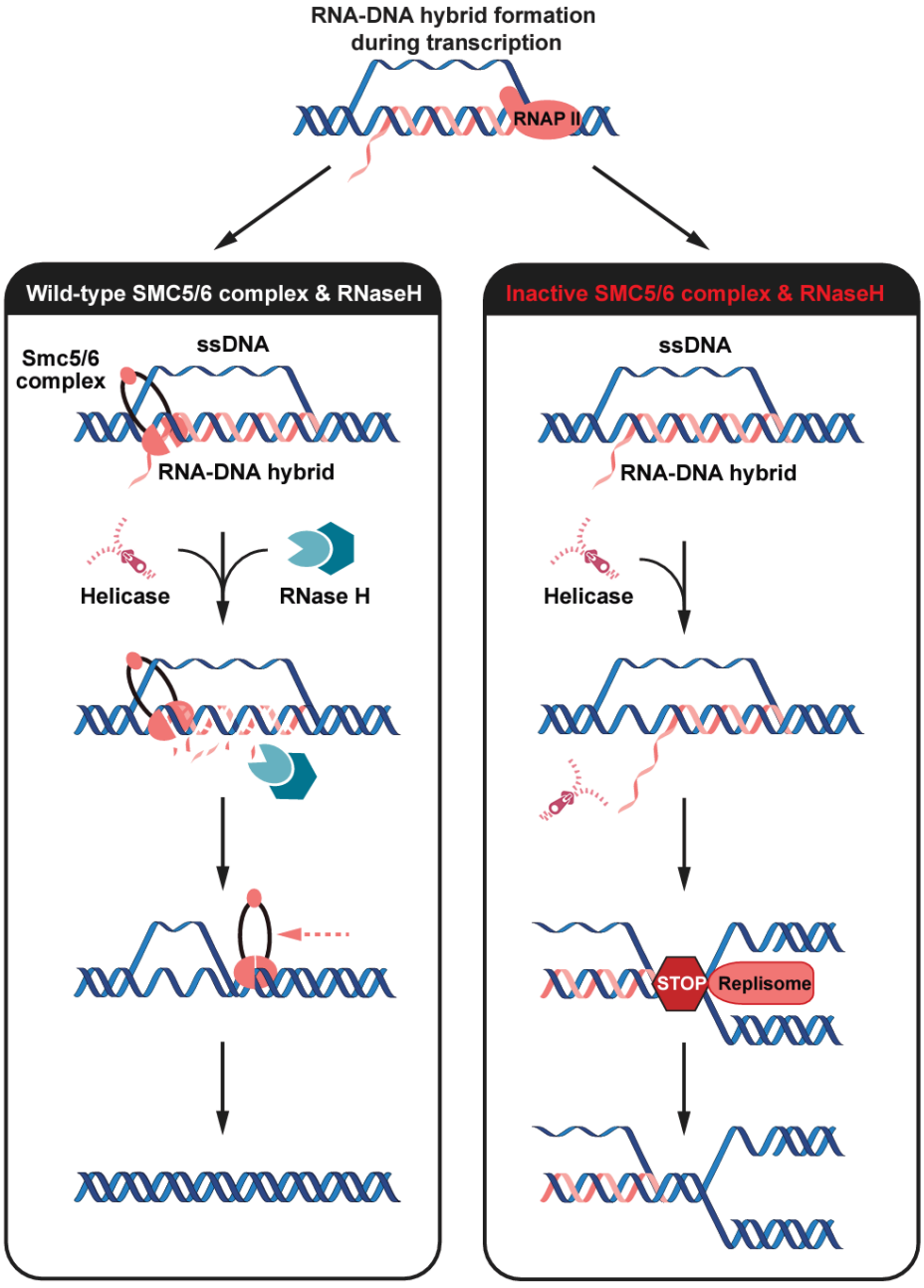


Figure 3. 7: Proposed mode of action for the Smc5/6 complex during R-loop removal from chromosomes. A nascent RNA transcript synthesized during gene transcription invades separated DNA strands and forms a stable interaction with its complementary DNA stand. The Smc5/6 complex then recognizes the R-loop and associates stably with the RNA-DNA hybrid structure. RNase H2 catalytic activity is stimulated in presence of the Smc5/6 complex. Effective removal of the RNA moiety from the R-loop promotes efficient reannealing of complementary ssDNA (left panel). The process prevents re-invasion of separated DNA strands by a new RNA transcript. In

the absence of RNase H and Smc5/6 complex, the stabilized R-loop will often cause replication stress and DNA double strand breaks (right panel). Figure created using Biorender and Adobe Illustrator.

Our genetic enhancement results obtained with double mutants must be interpreted carefully because they involve conditional/hypomorphic alleles of Smc5/6 complex components. Synthetic or enhancement phenotypes involving non-null alleles reflect the contribution of two mutations to the same cellular process, but not necessarily or exclusively in the same molecular pathway (*i.e*., interactions "within pathways" and "between pathways"; reviewed in[75–77]. As such, the exacerbation of the DNA damage sensitivity of RNase H mutants by temperature-sensitive alleles of the Smc5/6 complex may be the consequence of their effects on RNA-DNA hybrid removal (*i.e*., thus reflecting a "within pathway" contribution relative to RNase H1/H2) and their roles in additional biochemical pathways distinct from RNA-DNA hybrid degradation but still relevant to R-loop detoxification. We favor a model where the Smc5/6 complex acts at two distinct levels – within and between pathways– in the cellular response to R-loop formation (Figure 3.7). First, Smc5/6 contributes to RNase H-dependent removal of RNA-DNA hybrids from genomic DNA, as shown in Figures 3.1 and 3.2. In the execution of this function, Smc5/6 role is substantial but not as extensive as that of RNase H enzymes (more on this below). Second, the Smc5/6 complex plays an important role in the maintenance of DNA replication fork stability, as previously established[45]. This role is crucial for cellular fitness in the presence of elevated R-loop levels because these structures often disrupt replication fork progression and can lead to fork collapse[25,27,70,78]. In this context, losing both RNase H and Smc5/6 complex activities will have consequences well beyond those observed in individual mutants because it will increase R-loop

formation in a context where cells have lost the ability to cope with stress at replication forks. This "dual hit" will render cells hypersensitive to R-loops, thus explaining the synthetic effects of combining *smc5/6* and *rnase H* mutations. Interestingly, the dual hit scenario leading to additive phenotypes appears to be the prevalent paradigm observed with effectors of RNA-DNA hybrid metabolism. For instance, past genetic interaction studies have shown that combining *sen1-1* or *sen1-3* alleles with *rnh1Δ rnh201Δ* mutations leads to synthetic lethality[63]. Likewise, inactivation of the THO complex induces a substantial increase in RNA-DNA hybrids in cells defective in RNase H activity[79], similar to our observation with *nse4-4 rnh1Δ rnh201Δ* and *smc6-9 rnh1Δ rnh201Δ* mutant strains (Figure 3.2). This pattern of synthetic enhancements when inactivating effectors of RNA-DNA hybrid metabolism is consistent with multiple mechanisms acting independently to promote removal of toxic or unscheduled R-loops from eukaryotic genomes.

How might the Smc5/6 complex association with R-loops promote their degradation *in vivo*? Hints of a potential mechanism of action come from the observation that one of the enzymes responsible for R-loop degradation, RNase H1, shows little enzymatic activity under basal conditions and requires stimulation by ancillary factors, such as RPA, to achieve maximal R-loop degradation[80]. It therefore seems plausible that RNase H2 might also require the assistance of a separate stimulatory factor to achieve maximal catalytic efficiency. Testing this notion in a reconstituted R-loop degradation assay confirmed the model that RNase H2 activity can be stimulated effectively, and in a dose-dependent manner by the Smc5/6 complex. The two RNase H enzymes differ, however, in that RNase H1 shows very little RNA degradation activity in the absence of RPA[80], whereas RNase H2 is moderately active as an R-loop degrading enzyme at basal

state (Figure 3.6 and 3.S4B-D). The implication for the Smc5/6 complex is that it is probably not required to stimulate RNase H2 activity in all contexts *in vivo* but is likely more important in challenging environments where R-loops are highly abundant or otherwise difficult to degrade effectively by RNase H2 alone. This interpretation dovetails nicely with the synthetic interaction profiles we observed when combining Smc5/6 complex mutations with mutants that increase R-loop formation at selected genomic locations (Figure 3.4). Taken together, our genetic analyses indicate that the repertoire of genomic lesions that are substrates for RNase H2 and the Smc5/6 complex *in vivo* is not fully overlapping.

An RNase H2 stimulatory role for the Smc5/6 complex is compelling because it provides a cellular capacity/buffer to address substantial fluctuations in the total load of RNA-DNA hybrids produced under physiological and non-physiological conditions[67,70]. Consequently, reducing the total load of RNA-DNA hybrids or altering its sources of origin is expected to modify the requirement for the Smc5/6 complex in R-loop metabolism[41]. How might Smc5/6 stimulate RNase H2 enzymatic activity? This is a question for a future study, but it has not escaped our attention that RNase H1 is strongly stimulated by a ssDNA binding protein, RPA[80] , a biochemical property also encoded in the Smc5/6 complex[42,81–83]. Separate from this possibility, the Smc5/6 complex plays a vital role in promoting RPA binding and maintenance at ssDNA during homologous recombination[47], a function that could indirectly stimulate RNase H1 activity and R-loop repair[80] . Addressing these possibilities will require the identification of mutations abrogating the ssDNA binding activity of the Smc5/6 complex, a difficult feat for a holoenzyme known to associate with DNA through multiple different binding modes and domains (*i.e.*, topological and electrostatic;[42,83]).

Up to now, the Smc5/6 complex has been thought of primarily as a genome stability factor associated with the repair of DSBs, recovery of stalled replication forks, telomeric length maintenance, and virus restriction[45]. While it is not evident why the Smc5/6 complex would be involved in such a diverse and loosely connected group of cellular functions, it is nevertheless clear that failure to execute these functions generates toxic recombination intermediates *in vivo*. Under normal circumstances, the Smc5/6 complex binds to ssDNA intermediates and branched/structured substrates at DNA repair sites and the association of the complex to these DNA intermediates provides a platform for repair factors to resolve toxic DNA lesions[42,46,47]. It is interesting to note that most of the nuclear processes outlined above involve the formation of RNA-DNA hybrids in one form or another[27,70]. As such, the function we uncovered for the Smc5/6 complex in RNA-DNA metabolism may be a unifying role that explains its involvement in such a diverse repertoire of cellular functions. More work will be required to test this exciting possibility.

In conclusion, our work demonstrates a direct and active involvement of the Smc5/6 complex in the removal of R-loops from eukaryotic chromosomes. We showed the Smc5/6 complex binds strongly to RNA-DNA hybrid structures formed during active gene transcription and telomere length regulation, and subsequently promotes the removal of these toxic structures via the stimulation of RNase H2 enzymatic activity. This work uncovered a previously unanticipated contribution of the Smc5/6 complex in genome stability with important ramifications for the health and disease of all eukaryotic organisms.

Acknowledgments

We thank Sarah Kinkley, Malika Saint and members of the D'Amours laboratory for their comments on the manuscript. We also thank Marco Muzi-Falconi (Università degli Studi di Milano) for providing us with RNase H mutant strains, Pascal Chartrand (Université de Montréal) for sharing the *rat1-1* mutant and Robert J. Crouch (NIH) for providing the plasmids used for the purification of RNase H2. This work was supported by a CIHR grant to DD (FDN–167265). DD is also supported by a Canada Research Chair in Chromatin Dynamics & Genome Architecture (CRC-2017-00064).

Author contributions

SR and DD conceived and designed the experiments; SR created yeast strains, assessed their viability and DNA damage sensitivity; SR performed all the microscopy and DRIP experiments. HA performed all the biochemical experiments; SI performed the RNaseH2-SMC5/6 complex binding assay; SR, HA and DD analyzed the data and prepared figures; SR, HA and DD wrote the paper.

Declaration of Interests

The authors declare no competing interests.

3.5 METHODS

Yeast strains and cell viability assay

All yeast strains used in this study are derivatives of strain K699/K700. The complete list of the

relevant genotypes of the strains used in the study are provided in Table 3.S1. Yeast growth conditions, procedures for genetic analysis and creation of strains carrying relevant mutations was performed as previously described in [42]. To create double mutant strains, haploid mutants carrying the specified alleles were mated to produce heterozygous diploid yeasts. Sporulation and dissection of diploid strains was subsequently performed at 23°C. For cell viability assay, performed under conditions of DNA damage or replication stress, yeast were grown on solid medium containing MMS, 4-NQO, and HU at different temperatures. Specifically, 5-fold dilution series of wild-type and mutant yeast cultures (first spot on the left side of the plate corresponds to a culture at OD_{600} of 0.2) were spotted on solid YPD (yeast extract, peptone, 2% glucose) with or without the presence of DNA-damaging agent and grown in temperature-controlled incubators for 28-72 hours before scanning the plates in a scanner[42]. Strains expressing RNase H1 were generated by integrating *YIplac204::Pgal1::RNH1* at the *TRP1* locus and integration was confirmed by PCR screening. To create strains expressing the AID enzyme, we transformed relevant strains tagged with GFP at the C-terminus of Rad52 with *pESC-LEU-HsAIDSc* plasmid (Addgene plasmid #60810)[84]. A complete list of plasmids used in this study is also provided in Table 3.S2.

Chromatin spread immunolabelling with the S9.6 antibody

Chromatin spreads were performed as described previously[85] with some minor modifications. Briefly, Cells from mid-log phase grown in YPD at 23°C were collected and washed in 1 mL ZK buffer (25mM Tris, pH 7.5, 0.8M KCl) and were resuspended in ZK buffer supplemented with 1M DTT. Spheroplasting of cells was achieved by the addition of 5µl of Zymolyase (20 mg/ml) and

incubation at 30°C with gentle rotation. Subsequently, spheroplast cells were centrifuged (2000 rpm/5min) and resuspended in MES/Sorbitol buffer (0.1M MES pH 6.5, 0.5mM $MgCl_2$, 1mM EDTA, 1M sorbitol). Next, cells were added to the glass slide (Corning) and immediately were fixed and lysed by addition of a fixative solution (3% paraformaldehyde in 3.4% sucrose) and 1% NP-40 substitute solution. Cells were spread using a plastic pipette rolled from one end of the slide to the other end. Slides with chromatin spreads were dried overnight. The next day, slides were washed with 1x TBS (Tris-buffered saline) for 10 min and blocked for 15 min with 5% BSA (Bovine serum albumin) in 1x TBS. Chromatin spreads were incubated with mouse monoclonal antibody S9.6 (MABE1095) (1:250 dilution) for overnight followed by Cy3-conjugated goat anti mouse antibody (Jackson Laboratories, #115-165-003) for 2 hours. Nuclei were counterstained with 50 µl of VectaShield (Vector Laboratories, CA) plus 1x DAPI (4',6- diamidino-2-phenylindole) and sealed with nail polish. Images were acquired using Nikon Eclipse Ti2 inverted microscopy with an oil immersion 100x objective. For each replicate (n>3), about 150 nuclei were visualized and manually counted to obtain the fraction with detectable RNA-DNA hybrid foci using the 3D measurement module of the NIS-Elements software (Nikon Instruments Inc.).

RNA-DNA hybrid immunoprecipitation followed by qPCR

Mid-log cultures grown in YPD at 23°C were collected. RNA-DNA hybrids were processed and analyzed as described in[59]. Real-time quantitative PCR was performed at the indicated regions using the SsoAdvanced SYBR Green PCR Master Mix (Bio-Rad) with a CFX384 Real-Time PCR System (C-1000 Touch Thermal Cycler). Data was analyzed using the CFX Maestro Bio-Rad software and the relative abundance of RNA-DNA hybrid immunoprecipitated in each region was

normalized to the signal obtained in the inputs. Average and standard error of at least three independent experiments are shown.

Purification of the SMC5/6 complex

The human SMC5/6 core complex was purified using a triple affinity purification approach followed by size exclusion chromatography, as described by[42] with minor modifications. 35 L of a *S. cerevisiae* strain overexpressing the core complex was cultured under optimal growth conditions in a bioreactor (Techfors-S-42L) to an OD_{600} of 0.7 – 1.0. Protein expression was induced by addition of 2% galactose and cells were grown further for 16 hours at 18°C. Briefly, yeast pellets were resuspended in 200 ml buffer N (50 mM K_2HPO_4 / KH_2PO_4 pH8, 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol, 0.5% Triton X-100, 2 mM β-Mercaptoethanol) supplemented with 20 mM imidazole and protease inhibitors (E64, Pepstatin A, 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride [AESBF]). Yeast popcorn is made by the dropwise freezing of the cell suspension in liquid nitrogen. The popcorns were further lysed two cycles in a freezer mill. The lysates were resuspended in 1L of buffer N and passed through a high-pressure homogenizer (Avestin EmulsiFlex-C3) at an operating pressure of 25000 psi. The final lysate was centrifuged at 24000 rpm for 45 min at 4°C. The soluble lysates were passed through a column packed with Nickel-NTA resin at a flow rate of 5ml/minute for slower binding. The unbound fractions from the first purification column were loaded to a second Ni-NTA column for a second-round binding to maximise yield of purified protein. Both the columns were washed with 10 column volumes (CV) of buffer N supplemented with 60 mM imidazole. Complex was eluted with buffer SB (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol, 0.5% Tween 20, 2mM βME)

supplemented with 500 mM imidazole. The combined fractions from both the columns were loaded into a StrepTractin XT 5mL column using an AKTA prime FPLC purification system. The column was programed to wash with 10 CV of buffer SB supplemented with 0.5% Triton X-100 and eluted with 5 CV of buffer GB (25 mM K_2HPO_4/KH_2PO_4 pH8, 500 mM NaCl, 10% glycerol and 2 mM βME) supplemented with 50 mM biotin. The elution was mixed and incubated with 5 ml of pre-equilibrated Glutathione S-transferase (GST)-Sepharose resin, in GST binding buffer GB (20 mM K_2HPO_4/KH_2PO_4 pH8, 200 mM NaCl, 10% glycerol and 2 mM DTT) for 2h at 4°C. The resin was washed with 10 CV of buffer GEB (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol and 2 mM βME). The SMC5/6 complex was eluted with 5 CV of buffer GEB supplemented with 25 mM of reduced Glutathione. Linker, poly-histidine, Streptavidin-tag II and GST tags were cleaved by an overnight digestion with 1 mg of TEV protease per 4 mg/mL of fusion protein. Digestion was carried out in GEB buffer supplemented with 1 mM DTT. Digestion product was loaded into a Superose 6 10/300 size exclusion chromatography column in GF buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol and 2 mM βME) in order to remove the cleaved tags, digested linker and TEV protease. Elution fractions containing highly purified and stoichiometric complex were concentrated, quantified, snap frozen and stored at −80°C.

Purification of active RNase H2 enzyme complex

A polycistronic vector allowing the co-expression of all the subunits of human RNase H2 (*pET-hH2ABC;* was obtained from Robert J. Crouch (NIH). All the subunits of the holoenzyme –namely, RNase H2A, H2B, and H2C– are expressed from this vector as N-terminal hexahistidine fusion proteins. For purification, *Escherichia coli* BL21 was transformed with *pET-hH2ABC* and 6L of

culture was grown at 37°C to an OD_{600} of 0.4 - 0.6 before being induced with 0.3mM of IPTG. The culture was grown further at 18°C for 16h after induction. Bacterial pellets were resuspended in 100 ml buffer N (50 mM K_2HPO_4 / KH_2PO_4 pH8, 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol, 0.5% Triton X-100, 2 mM βME) supplemented with 20 mM imidazole and protease inhibitors (E64, Pepstatin A, AESBF). The lysate was passed twice through a high-pressure homogenizer (Avestin EmulsiFlex-C3) at an operating pressure of 15000 psi and centrifuged at 24000 rpm for 45 min at 4°C. The soluble lysates were passed through a column packed with Ni-NTA resin at a flow rate of 5ml/minute for slower binding. The column was washed with 10 column volumes (CV) of buffer N supplemented with 60 mM imidazole. Complex was eluted with buffer SB (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol, 0.5% tween 20, 2mM βME) supplemented with 500 mM imidazole. The eluted fractions were dialyzed against 20mM HEPES pH 7.6, 250mM NaCl, 10% glycerol, and 1mM DTT, quantified, snap frozen and stored at −80°C.

RNA-DNA hybrid probe synthesis

The DNA or RNA oligos are 5'-labeled with ATP-[γ^{32}P] (PerkinElmer Life Sciences) using T4 polynucleotide kinase (New England BioLabs). Radiolabeled oligos were then annealed to a complementary strand by heating to 95°C and slow cooling over a long period of time in PNK buffer (70 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT). Annealed substrates were separated from free ATP-[γ^{32}P] on an 8% native PAGE in Tris Borate/EDTA buffer at room temperature. The gel band corresponding to the annealed substrate was excised, purified, and finally eluted. The eluted substrates were quantified (nM) using scintillation counter.

Two types of R-loop substrates were synthesized. First, a genuine R-loop substrate was constructed by annealing [^{32}P]-labeled DNA strand 2 (DD 4264) with DNA strand 1 (DD 4263) and RNA strand (DD4265). Second, the R*-loop substrate was constructed by annealing [^{32}P]-labeled RNA strand (DD 4265) with DNA strand 1 (DD 4263) and DNA strand 2 (DD 4264). A D-loop substrate was also generated by annealing [^{32}P]-labelled DNA strand 2 with DNA strand 1 and DNA strand 3. A single [^{32}P]-labeled ssDNA strand (DD4264) was used as control for electrophoretic migration in gel.

The sequence of oligonucleotides used as *in vitro* substrates are: DNA strand 1(DD 4263; 1.5 µg/µl or100µM):5'GGGTGAACCTGCAGGTGGGCGGCTGCTCATCGTAGGTTAGTTGGTAGAATTCGGCAGCG TC-3' (61 mer); DNA strand 2 (DD 4264; 1.8 µg/µl or 100µM): 5'GACGCTGCCGAATTCTACCAGTGCCTTGC TAGGACATCTTTGCCCACCTGCAGGTTCACCC-3' (61 mer); RNA strand 1 (DD 4265; 0.5 µg/µl or 100µM): 5'-AAAGArUGrUCCrUAGCAAGGCAC-3' (21 mer); DNA strand 3 (DD 4266; 0.6 µg/µl or 100µM): 5'-AAAGATGTCCTAGCAAGGCAC-3' (21 mer).

In vitro RNA-DNA and dsDNA binding assays

Reactions containing 40 nM [^{32}P]-labelled oligonucleotides and the indicated concentrations of SMC5/6 complex were incubated in binding buffer A (25 mM MOPS [morpholinepropanesulfonic acid] pH 7.6, 60 mM KCl, 0.2% NP40, 2 mM DTT, 5 mM $MgCl_2$) in a total volume of 15µl. Reactions were incubated at 24°C for 10 min and loaded on an 6% acrylamide gel, electrophoresed at 150 volts for 240 min in 1X TBE buffer. Gels were then dried onto DE81 filter paper and visualized by autoradiography.

R-loop RNase H2 assay

RNase assays were performed in Buffer C (20mM HEPES [4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid] pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$, 0.5mM DTT). The R*-loop (40 nM) substrates were pre-incubated with SMC5/6 complex at the indicated concentration in buffer C for 10 min at 24°C followed by addition of RNase H2 complex for 7 min at the same reaction conditions. Reactions were deproteinized in one-fifth volume of stop buffer (Buffer A, 1% SDS, 5mM EDTA and 0.2 mg/ml proteinase K) for 15 min at 24°C. Reactions were loaded on an 8% acrylamide gel, electrophoresed at 150 volts for 150 min, dried onto DE81 filter paper and visualized by autoradiography.

AID-induced Rad52 foci assay by fluorescence microscopy

Detection of R-loops through AID-induced DNA damage and subsequent rise of Rad52 foci was performed as described in[60]. Strains were grown in minimal media followed by galactose induction for 2 hours to overexpress AID enzymes and collected. Cells were fixed in formaldehyde (10% in 0.1M KPO_4 pH6.4) for 30 minutes at room temperature and washed twice in 0.1M KPO_4 pH7.0 buffer. Staining of the nuclei was performed with DAPI at a final concentration of 2µg/ml in cells suspended in 0.1M KPO_4 pH7.0 buffer. For DAPI staining and Rad52-GFP visualization, images were acquired using Nikon Eclipse Ti2 inverted microscopy with an oil immersion 100x objective. For each replicate (n=3), about 100 cells were visualized and manually counted to obtain the fraction with detectable Rad52 foci using the 3D measurement module of the NIS-Elements software (Nikon Instruments Inc.).

In vitro RNaseH2 and SMC5/6 complex binding assay

Active RNase H2 enzyme and the SMC5/6 complex were purified as described above. The resulting complexes carry a Strep (3XSTII) tag on SMC6 and a 10xHis tag on NSMCE4. RNase H2 subunits were tagged with a 6xHis tag. For the pull-down procedure, streptactin XT resin (10µL per reaction) was washed twice with water and twice with binding buffer (50mM HEPES-NaOH [4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid, N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) – sodium hydroxide] pH 7.5, 150 mM NaCl, 5% glycerol, 1% BSA, 5mM MgCl2) before being incubated in binding buffer in a total volume of 300µL for 2hrs at 4°C with or without 0.8µM of the human SMC5/6 complex per reaction. The resins were then washed 4 times with washing buffer (50mM HEPES-NaOH pH 7.5, 150 mM NaCl, 5% glycerol, 0.2% NP40, 5mM MgCl2). For each condition, 10µL of streptactin XT resin with or without 1.95µM of human RNase H2 were incubated in a total volume of 300µL of binding buffer at 4°C overnight (~12hrs). The reactions were then washed 4 times with washing buffer. The pellet was resuspended in washing buffer and 4x sample buffer (90% 4x Laemmli Sample Buffer, 10% β-mercaptoethanol) at a ratio of 1:1. The reactions were then loaded on an SDS-PAGE gel in 1x MOPS buffer and ran for 2hrs at 120 Volts, before being analyzed by immunoblotting with 1:2500 dilution of anti-HIS antibody (Qiagen 34660).

Statistical analyses

Results presented in this study are representative examples of at least three independent experiments. All statistical analyses were performed using GraphPad Prism 7 (GraphPad Software

Inc). Sample size (n) and statistical tests performed in each experiment are described in the relevant figure legends. In all cases, *P* values expressed as **P* < 0.05, ***P* < 0.005, and ****P* < 0.0005 are considered significant.

Data availability

All source Data files have been provided for Figures 3.2-3.3, Figure 3.5-3.6, and Supplementary Figure 3.4 on the Dryad open repository site (DOI: 10.5061/dryad.3xsj3txpg). Reviewers can access the full data set through the following link:

https://datadryad.org/stash/share/4QICK1jufN1YT0elOow5UKTNIwhKEW-L0WsDZ0W1Dlc

3.6 REFERENCES

1. Negrini, S., Gorgoulis, V.G., and Halazonetis, T.D. (2010). Genomic instability — an evolving hallmark of cancer. Nat Rev Mol Cell Biol *11*, 220–228. 10.1038/nrm2858.

2. Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674. 10.1016/j.cell.2011.02.013.

3. Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1997). Genetic instability in colorectal cancers. Nature *386*, 623–627. 10.1038/386623a0.

4. Cifone, M.A., and Fidler, I.J. (1981). Increasing metastatic potential is associated with increasing genetic instability of clones isolated from murine neoplasms. Proceedings of the National Academy of Sciences *78*, 6949–6952. 10.1073/pnas.78.11.6949.

5. Tubbs, A., and Nussenzweig, A. (2017). Endogenous DNA Damage as a Source of Genomic Instability in Cancer. Cell *168*, 644–656. 10.1016/j.cell.2017.01.002.

6. Thada, V., and Greenberg, R.A. (2022). Unpaved roads: How the DNA damage response navigates endogenous genotoxins. DNA Repair *118*, 103383. 10.1016/j.dnarep.2022.103383.

7. Lindahl, T., and Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. Biochemistry *11*, 3610–3618. 10.1021/bi00769a018.

8. Chatzidoukaki, O., Stratigi, K., Goulielmaki, E., Niotis, G., Akalestou-Clocher, A., Gkirtzimanaki, K., Zafeiropoulos, A., Altmüller, J., Topalis, P., and Garinis, G.A. (2021). R-loops trigger the release of cytoplasmic ssDNAs leading to chronic inflammation upon DNA damage. Science Advances *7*, eabj5769. 10.1126/sciadv.abj5769.

9. Costantino, L., and Koshland, D. (2018). Genome-wide Map of R-Loop-Induced Damage Reveals How a Subset of R-Loops Contributes to Genomic Instability. Mol Cell *71*, 487-497.e3. 10.1016/j.molcel.2018.06.037.

10. De Magis, A., Manzo, S.G., Russo, M., Marinello, J., Morigi, R., Sordet, O., and Capranico, G. (2019). DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. Proceedings of the National Academy of Sciences *116*, 816–825. 10.1073/pnas.1810409116.

11. Stork, C.T., Bocek, M., Crossley, M.P., Sollier, J., Sanz, L.A., Chédin, F., Swigut, T., and Cimprich, K.A. (2016). Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. eLife *5*, e17548. 10.7554/eLife.17548.

12. Yu, K., Chedin, F., Hsieh, C.-L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol *4*, 442–451. 10.1038/ni919.

13. Roy, D., Yu, K., and Lieber, M.R. (2008). Mechanism of R-Loop Formation at Immunoglobulin Class Switch Sequences. Mol Cell Biol *28*, 50–60. 10.1128/MCB.01251-07.

14. Pohjoismäki, J.L.O., Holmes, J.B., Wood, S.R., Yang, M.-Y., Yasukawa, T., Reyes, A., Bailey, L.J., Cluett, T.J., Goffart, S., Willcox, S., et al. (2010). Mammalian Mitochondrial DNA Replication Intermediates Are Essentially Duplex but Contain Extensive Tracts of RNA/DNA Hybrid. Journal of Molecular Biology *397*, 1144–1155. 10.1016/j.jmb.2010.02.029.

15. Xu, B., and Clayton, D.A. (1996). RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers. The EMBO Journal *15*, 3135–3143. 10.1002/j.1460-2075.1996.tb00676.x.

16. Baker, T.A., and Kornberg, A. (1988). Transcriptional activation of initiation of replication from the E. coli chromosomal origin: An RNA-DNA hybrid near oriC. Cell *55*, 113–123. 10.1016/0092-8674(88)90014-1.

17. McLean, E.K., Nye, T.M., Lowder, F.C., and Simmons, L.A. (2022). The Impact of RNA-DNA Hybrids on Genome Integrity in Bacteria. Annual Review of Microbiology *76*, 461–480. 10.1146/annurev-micro-102521-014450.

18. Zhang, B., Luo, D., Li, Y., Perčulija, V., Chen, J., Lin, J., Ye, Y., and Ouyang, S. (2021). Mechanistic insights into the R-loop formation and cleavage in CRISPR-Cas12i1. Nat Commun *12*, 3476. 10.1038/s41467-021-23876-5.

19. Xiao, Y., Luo, M., Hayes, R.P., Kim, J., Ng, S., Ding, F., Liao, M., and Ke, A. (2017). Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. Cell *170*, 48-60.e11. 10.1016/j.cell.2017.06.012.

20. Sidorenkov, I., Komissarova, N., and Kashlev, M. (1998). Crucial role of the RNA:DNA hybrid in the processivity of transcription. Mol Cell *2*, 55–64. 10.1016/s1097-2765(00)80113-6.

21. Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell *42*, 794–805. 10.1016/j.molcel.2011.04.026.

22. Nudler, E., Mustaev, A., Goldfarb, A., and Lukhtanov, E. (1997). The RNA–DNA Hybrid Maintains the Register of Transcription by Preventing Backtracking of RNA Polymerase. Cell *89*, 33–41. 10.1016/S0092-8674(00)80180-4.

23. Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K., and Luke, B. (2013). Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. Nat Struct Mol Biol *20*, 1199–1205. 10.1038/nsmb.2662.

24. Luke, B., Panza, A., Redon, S., Iglesias, N., Li, Z., and Lingner, J. (2008). The Rat1p 5' to 3' exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in Saccharomyces cerevisiae. Mol Cell *32*, 465–477. 10.1016/j.molcel.2008.10.019.

25. Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019). R-Loops as Cellular Regulators and Genomic Threats. Molecular Cell *73*, 398–411. 10.1016/j.molcel.2019.01.024.

26. García-Muse, T., and Aguilera, A. (2019). R Loops: From Physiological to Pathological Roles. Cell *179*, 604–618. 10.1016/j.cell.2019.08.055.

27. Brambati, A., Zardoni, L., Nardini, E., Pellicioli, A., and Liberi, G. (2020). The dark side of RNA:DNA hybrids. Mutat Res Rev Mutat Res *784*, 108300. 10.1016/j.mrrev.2020.108300.

28. Niehrs, C., and Luke, B. (2020). Regulatory R-loops as facilitators of gene expression and genome stability. Nat Rev Mol Cell Biol *21*, 167–178. 10.1038/s41580-019-0206-3.

29. Zardoni, L., Nardini, E., Brambati, A., Lucca, C., Choudhary, R., Loperfido, F., Sabbioneda, S., and Liberi, G. (2021). Elongating RNA polymerase II and RNA:DNA hybrids hinder fork progression and gene expression at sites of head-on replication-transcription collisions. Nucleic Acids Research *49*, 12769–12784. 10.1093/nar/gkab1146.

30. Aguilera, A., and García-Muse, T. (2012). R Loops: From Transcription Byproducts to Threats to Genome Stability. Molecular Cell *46*, 115–124. 10.1016/j.molcel.2012.04.009.

31. Santos-Pereira, J.M., and Aguilera, A. (2015). R loops: new modulators of genome dynamics and function. Nat Rev Genet *16*, 583–597. 10.1038/nrg3961.

32. Hamperl, S., Bocek, M.J., Saldivar, J.C., Swigut, T., and Cimprich, K.A. (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. Cell *170*, 774-786.e19. 10.1016/j.cell.2017.07.043.

33. Kumar, C., Batra, S., Griffith, J.D., and Remus, D. (2021). The interplay of RNA:DNA hybrid structure and G-quadruplexes determines the outcome of R-loop-replisome collisions. eLife *10*, e72286. 10.7554/eLife.72286.

34. Kim, S., Shin, W.H., Kang, Y., Kim, H., and Lee, J.Y. (2024). Direct visualization of replication and R-loop collision using single-molecule imaging. Nucleic Acids Research *52*, 259–273. 10.1093/nar/gkad1101.

35. Freudenreich, C.H. (2018). R-loops: targets for nuclease cleavage and repeat instability. Curr Genet *64*, 789–794. 10.1007/s00294-018-0806-z.

36. Miglietta, G., Russo, M., and Capranico, G. (2020). G-quadruplex–R-loop interactions and the mechanism of anticancer G-quadruplex binders. Nucleic Acids Res *48*, 11942–11957. 10.1093/nar/gkaa944.

37. Richard, P., and Manley, J.L. (2017). R Loops and Links to Human Disease. J Mol Biol *429*, 3168–3180. 10.1016/j.jmb.2016.08.031.

38. Lazzaro, F., Novarina, D., Amara, F., Watt, D.L., Stone, J.E., Costanzo, V., Burgers, P.M., Kunkel, T.A., Plevani, P., and Muzi-Falconi, M. (2012). RNase H and postreplication repair protect cells from ribonucleotides incorporated in DNA. Mol Cell *45*, 99–110. 10.1016/j.molcel.2011.12.019.

39. Lockhart, A., Pires, V.B., Bento, F., Kellner, V., Luke-Glaser, S., Yakoub, G., Ulrich, H.D., and Luke, B. (2019). RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids. Cell Reports *29*, 2890-2900.e5. 10.1016/j.celrep.2019.10.108.

40. Camino, L.P., Dutta, A., Barroso, S., Pérez-Calero, C., Katz, J.N., García-Rubio, M., Sung, P., Gómez-González, B., and Aguilera, A. (2023). DICER ribonuclease removes harmful R-loops. Molecular Cell *83*, 3707-3719.e5. 10.1016/j.molcel.2023.09.021.

41. Lafuente-Barquero, J., Luke-Glaser, S., Graf, M., Silva, S., Gómez-González, B., Lockhart, A., Lisby, M., Aguilera, A., and Luke, B. (2017). The Smc5/6 complex regulates the yeast Mph1 helicase at RNA-DNA hybrid-mediated DNA damage. PLOS Genetics *13*, e1007136. 10.1371/journal.pgen.1007136.

42. Serrano, D., Cordero, G., Kawamura, R., Sverzhinsky, A., Sarker, M., Roy, S., Malo, C., Pascal, J.M., Marko, J.F., and D'Amours, D. (2020). The Smc5/6 Core Complex Is a Structure-Specific DNA Binding and Compacting Machine. Mol Cell *80*, 1025-1038.e5. 10.1016/j.molcel.2020.11.011.

43. Girasol, M.J., Briggs, E.M., Marques, C.A., Batista, J.M., Beraldi, D., Burchmore, R., Lemgruber, L., and McCulloch, R. (2023). Immunoprecipitation of RNA–DNA hybrid interacting proteins in Trypanosoma brucei reveals conserved and novel activities, including in the control of surface antigen expression needed for immune evasion by antigenic variation. Nucleic Acids Research, gkad836. 10.1093/nar/gkad836.

44. Penzo, A., Dubarry, M., Brocas, C., Zheng, M., Mangione, R.M., Rougemaille, M., Goncalves, C., Lautier, O., Libri, D., Simon, M.-N., et al. (2023). A R-loop sensing pathway mediates the relocation of transcribed genes to nuclear pore complexes. Nat Commun *14*, 5606. 10.1038/s41467-023-41345-z.

45. Peng, X.P., and Zhao, X. (2023). The multi-functional Smc5/6 complex in genome protection and disease. Nat Struct Mol Biol *30*, 724–734. 10.1038/s41594-023-01015-6.

46. Gutierrez-Escribano, P., Hormeño, S., Madariaga-Marcos, J., Solé-Soler, R., O'Reilly, F.J., Morris, K., Aicart-Ramos, C., Aramayo, R., Montoya, A., Kramer, H., et al. (2020). Purified

Smc5/6 Complex Exhibits DNA Substrate Recognition and Compaction. Mol Cell *80*, 1039-1054.e6. 10.1016/j.molcel.2020.11.012.

47. Tanasie, N.-L., Gutiérrez-Escribano, P., Jaklin, S., Aragon, L., and Stigler, J. (2022). Stabilization of DNA fork junctions by Smc5/6 complexes revealed by single-molecule imaging. Cell Rep *41*, 111778. 10.1016/j.celrep.2022.111778.

48. Pradhan, B., Kanno, T., Umeda Igarashi, M., Loke, M.S., Baaske, M.D., Wong, J.S.K., Jeppsson, K., Björkegren, C., and Kim, E. (2023). The Smc5/6 complex is a DNA loop-extruding motor. Nature *616*, 843–848. 10.1038/s41586-023-05963-3.

49. Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D., et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. Science *353*, aaf1420. 10.1126/science.aaf1420.

50. Styles, E.B., Founk, K.J., Zamparo, L.A., Sing, T.L., Altintas, D., Ribeyre, C., Ribaud, V., Rougemont, J., Mayhew, D., Costanzo, M., et al. (2016). Exploring Quantitative Yeast Phenomics with Single-Cell Analysis of DNA Damage Foci. Cell Syst *3*, 264-277.e10. 10.1016/j.cels.2016.08.008.

51. Kuzmin, E., VanderSluis, B., Wang, W., Tan, G., Deshpande, R., Chen, Y., Usaj, M., Balint, A., Mattiazzi Usaj, M., van Leeuwen, J., et al. (2018). Systematic analysis of complex genetic interactions. Science *360*, eaao1729. 10.1126/science.aao1729.

52. Chang, E.Y.-C., Tsai, S., Aristizabal, M.J., Wells, J.P., Coulombe, Y., Busatto, F.F., Chan, Y.A., Kumar, A., Dan Zhu, Y., Wang, A.Y.-H., et al. (2019). MRE11-RAD50-NBS1 promotes Fanconi Anemia R-loop suppression at transcription-replication conflicts. Nat Commun *10*, 4265. 10.1038/s41467-019-12271-w.

53. Ben-Aroya, S., Coombes, C., Kwok, T., O'Donnell, K.A., Boeke, J.D., and Hieter, P. (2008). Toward a comprehensive temperature-sensitive mutant repository of the essential genes of Saccharomyces cerevisiae. Mol Cell *30*, 248–258. 10.1016/j.molcel.2008.02.021.

54. Hwang, J.-Y., Smith, S., Ceschia, A., Torres-Rosell, J., Aragon, L., and Myung, K. (2008). Smc5-Smc6 complex suppresses gross chromosomal rearrangements mediated by break-induced replications. DNA Repair (Amst) *7*, 1426–1436. 10.1016/j.dnarep.2008.05.006.

55. Nowotny, M., Gaidamakov, S.A., Ghirlando, R., Cerritelli, S.M., Crouch, R.J., and Yang, W. (2007). Structure of Human RNase H1 Complexed with an RNA/DNA Hybrid: Insight into HIV Reverse Transcription. Molecular Cell *28*, 264–276. 10.1016/j.molcel.2007.08.015.

56. Figiel, M., Chon, H., Cerritelli, S.M., Cybulska, M., Crouch, R.J., and Nowotny, M. (2011). The Structural and Biochemical Characterization of Human RNase H2 Complex Reveals the Molecular Basis for Substrate Recognition and Aicardi-Goutières Syndrome Defects*. Journal of Biological Chemistry *286*, 10540–10550. 10.1074/jbc.M110.181974.

57. Hallett, S.T., Campbell Harry, I., Schellenberger, P., Zhou, L., Cronin, N.B., Baxter, J., Etheridge, T.J., Murray, J.M., and Oliver, A.W. (2022). Cryo-EM structure of the Smc5/6 holo-complex. Nucleic Acids Res *50*, 9505–9520. 10.1093/nar/gkac692.

58. Bou-Nader, C., Bothra, A., Garboczi, D.N., Leppla, S.H., and Zhang, J. (2022). Structural basis of R-loop recognition by the S9.6 monoclonal antibody. Nat Commun *13*, 1641. 10.1038/s41467-022-29187-7.

59. El Hage, A., and Tollervey, D. (2018). Immunoprecipitation of RNA:DNA Hybrids from Budding Yeast. Methods Mol Biol *1703*, 109–129. 10.1007/978-1-4939-7459-7_8.

60. Cañas, J.C., Aguilera, A., and Gómez-González, B. (2022). Detection of R-Loops by In Vivo and In Vitro Cytosine Deamination in Saccharomyces cerevisiae. Methods Mol Biol *2528*, 39–53. 10.1007/978-1-0716-2477-7_4.

61. Luna, R., Rondón, A.G., Pérez-Calero, C., Salas-Armenteros, I., and Aguilera, A. (2019). The THO Complex as a Paradigm for the Prevention of Cotranscriptional R-Loops. Cold Spring Harb Symp Quant Biol *84*, 105–114. 10.1101/sqb.2019.84.039594.

62. Mischo, H.E., Gómez-González, B., Grzechnik, P., Rondón, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. Mol Cell *41*, 21–32. 10.1016/j.molcel.2010.12.007.

63. Appanah, R., Lones, E.C., Aiello, U., Libri, D., and De Piccoli, G. (2020). Sen1 Is Recruited to Replication Forks via Ctf4 and Mrc1 and Promotes Genome Stability. Cell Rep *30*, 2094-2105.e9. 10.1016/j.celrep.2020.01.087.

64. Nick McElhinny, S.A., Kumar, D., Clark, A.B., Watt, D.L., Watts, B.E., Lundström, E.-B., Johansson, E., Chabes, A., and Kunkel, T.A. (2010). Genome instability due to ribonucleotide incorporation into DNA. Nat Chem Biol *6*, 774–781. 10.1038/nchembio.424.

65. Zimmer, A.D., and Koshland, D. (2016). Differential roles of the RNases H in preventing chromosome instability. Proceedings of the National Academy of Sciences *113*, 12220–12225. 10.1073/pnas.1613448113.

66. Chon, H., Vassilev, A., DePamphilis, M.L., Zhao, Y., Zhang, J., Burgers, P.M., Crouch, R.J., and Cerritelli, S.M. (2009). Contributions of the two accessory subunits, RNASEH2B and RNASEH2C, to the activity and properties of the human RNase H2 complex. Nucleic Acids Res *37*, 96–110. 10.1093/nar/gkn913.

67. Mackay, R.P., Xu, Q., and Weinberger, P.M. (2020). R-Loop Physiology and Pathology: A Brief Review. DNA and Cell Biology *39*, 1914–1925. 10.1089/dna.2020.5906.

68. Crossley, M.P., Song, C., Bocek, M.J., Choi, J.-H., Kousorous, J., Sathirachinda, A., Lin, C., Brickner, J.R., Bai, G., Lans, H., et al. (2023). R-loop-derived cytoplasmic RNA-DNA hybrids activate an immune response. Nature *613*, 187–194. 10.1038/s41586-022-05545-9.

69. Moradi-Fard, S., Sarthi, J., Tittel-Elmer, M., Lalonde, M., Cusanelli, E., Chartrand, P., and Cobb, J.A. (2016). Smc5/6 Is a Telomere-Associated Complex that Regulates Sir4 Binding and TPE. PLOS Genetics *12*, e1006268. 10.1371/journal.pgen.1006268.

70. Petermann, E., Lan, L., and Zou, L. (2022). Sources, resolution and physiological relevance of R-loops and RNA-DNA hybrids. Nat Rev Mol Cell Biol *23*, 521–540. 10.1038/s41580-022-00474-x.

71. Diman, A., Panis, G., Castrogiovanni, C., Prados, J., Baechler, B., and Strubin, M. (2023). Human Smc5/6 recognises transcription-generated positive DNA supercoils. Preprint at bioRxiv, 10.1101/2023.05.04.539344 10.1101/2023.05.04.539344.

72. Jeppsson, K., Pradhan, B., Sutani, T., Sakata, T., Igarashi, M.U., Berta, D.G., Kanno, T., Nakato, R., Shirahige, K., Kim, E., et al. (2023). Loop-extruding Smc5/6 organizes transcription-induced positive DNA supercoils. Preprint at bioRxiv, 10.1101/2023.06.20.545053 10.1101/2023.06.20.545053.

73. Jeppsson, K., Kanno, T., Shirahige, K., and Sjögren, C. (2014). The maintenance of chromosome structure: positioning and functioning of SMC complexes. Nature Reviews Molecular Cell Biology *15*, 601.

74. Pebernard, S., Schaffer, L., Campbell, D., Head, S.R., and Boddy, M.N. (2008). Localization of Smc5/6 to centromeres and telomeres requires heterochromatin and SUMO, respectively. EMBO J *27*, 3011–3023. 10.1038/emboj.2008.220.

75. Huang, L.S., and Sternberg, P.W. (1995). Genetic dissection of developmental pathways. Methods Cell Biol *48*, 97–122. 10.1016/s0091-679x(08)61385-0.

76. Boone, C., Bussey, H., and Andrews, B.J. (2007). Exploring genetic interactions and networks with yeast. Nat Rev Genet *8*, 437–449. 10.1038/nrg2085.

77. Roth, F.P., Lipshitz, H.D., and Andrews, B.J. (2009). Q&A: epistasis. J Biol *8*, 35. 10.1186/jbiol144.

78. Kemiha, S., Poli, J., Lin, Y.-L., Lengronne, A., and Pasero, P. (2021). Toxic R-loops: Cause or consequence of replication stress? DNA Repair (Amst) *107*, 103199. 10.1016/j.dnarep.2021.103199.

79. Yang, X., Zhai, B., Wang, S., Kong, X., Tan, Y., Liu, L., Yang, X., Tan, T., Zhang, S., and Zhang, L. (2021). RNA-DNA hybrids regulate meiotic recombination. Cell Reports *37*, 110097. 10.1016/j.celrep.2021.110097.

80. Nguyen, H.D., Yadav, T., Giri, S., Saez, B., Graubert, T.A., and Zou, L. (2017). Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1. Mol Cell *65*, 832-847.e4. 10.1016/j.molcel.2017.01.029.

81. Roy, M.-A., and D'Amours, D. (2011). DNA-binding properties of Smc6, a core component of the Smc5-6 DNA repair complex. Biochem Biophys Res Commun *416*, 80–85. 10.1016/j.bbrc.2011.10.149.

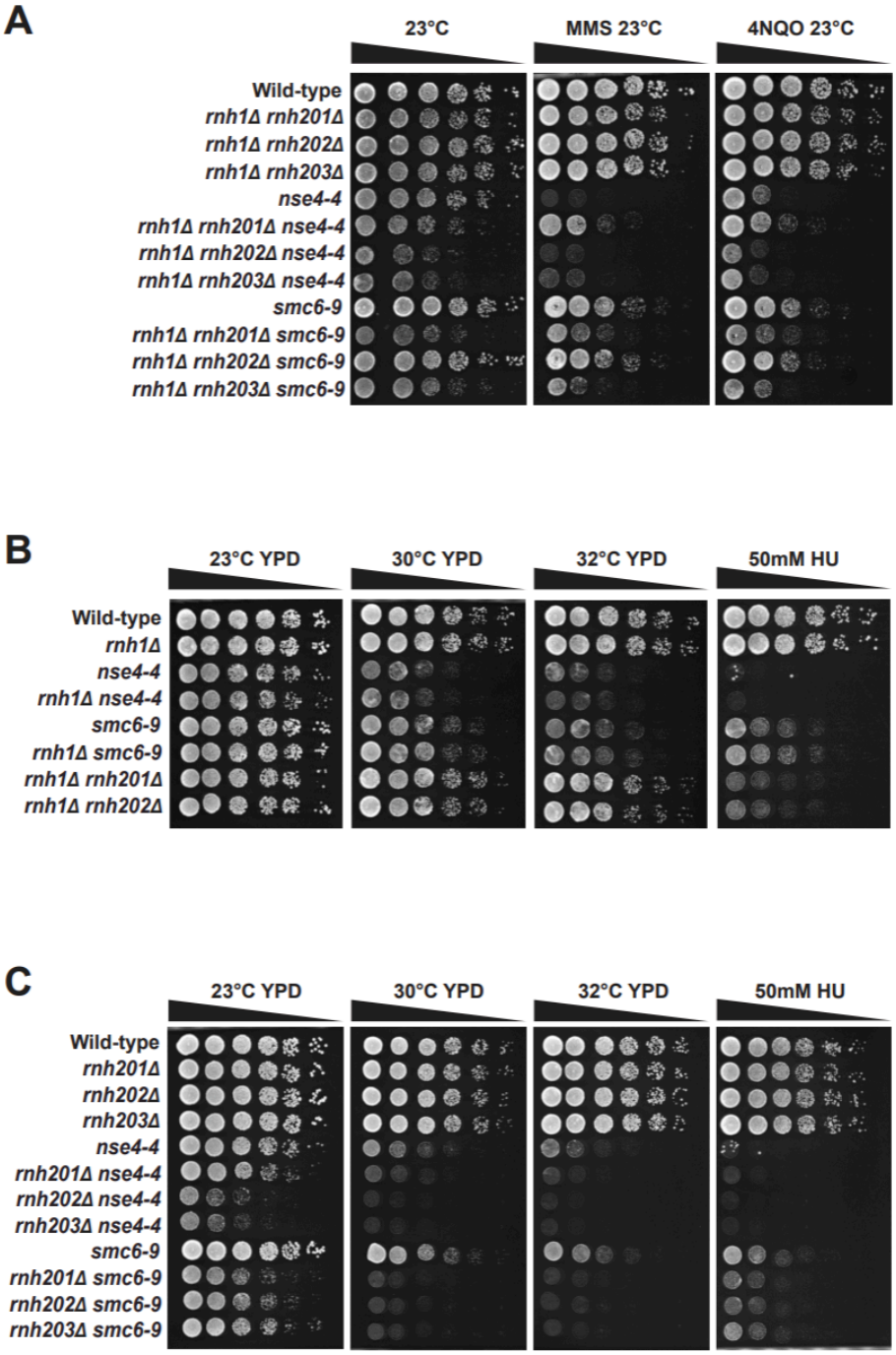
82. Roy, M.-A., Siddiqui, N., and D'Amours, D. (2011). Dynamic and selective DNA-binding activity of Smc5, a core component of the Smc5-Smc6 complex. Cell Cycle *10*, 690–700. 10.4161/cc.10.4.14860.

83. Roy, M.-A., Dhanaraman, T., and D'Amours, D. (2015). The Smc5-Smc6 heterodimer associates with DNA through several independent binding domains. Scientific Reports *5*, 9797.

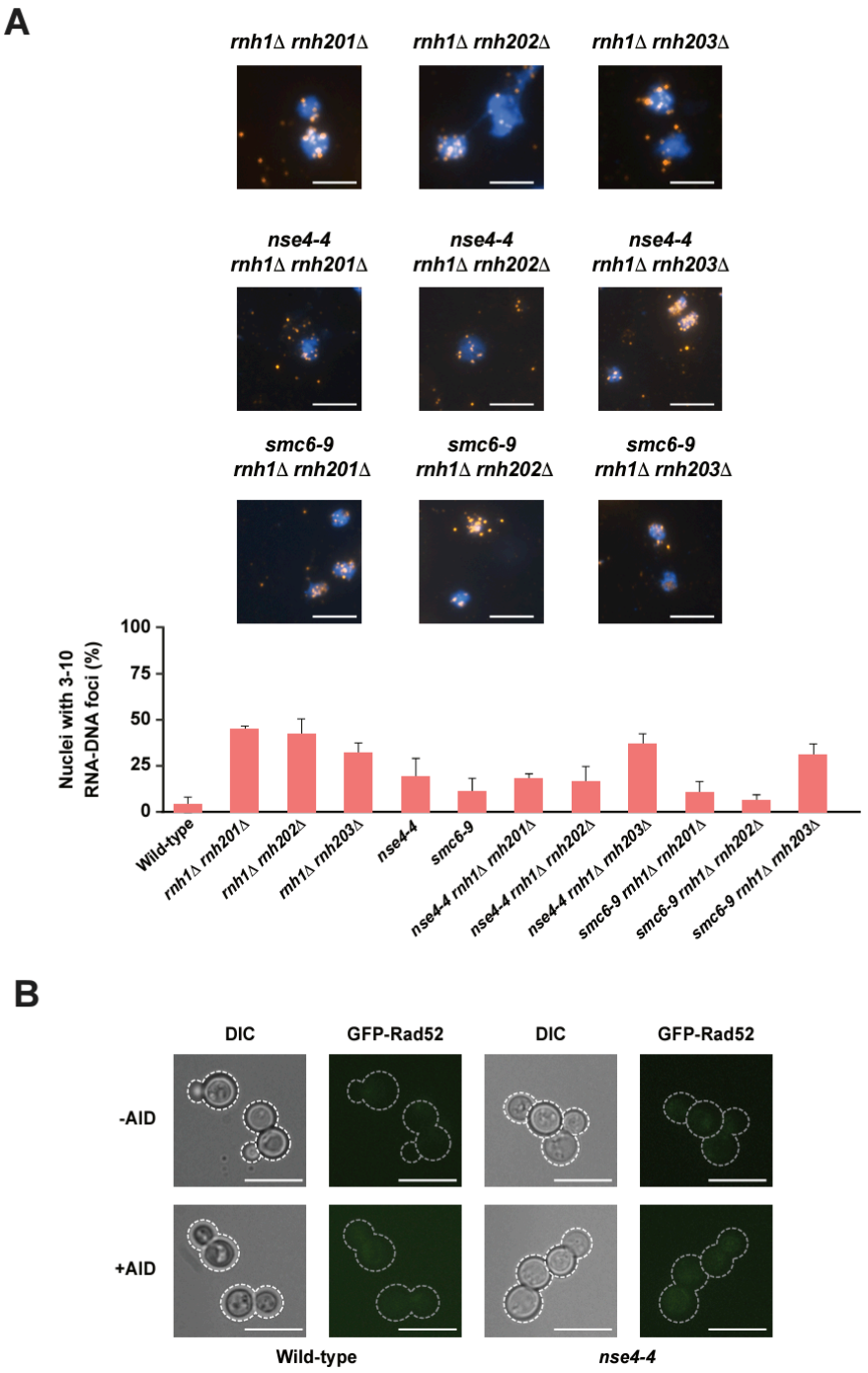
84. Mayorov, V.I., Rogozin, I.B., Adkison, L.R., Frahm, C., Kunkel, T.A., and Pavlov, Y.I. (2005). Expression of human AID in yeast induces mutations in context similar to the context of somatic hypermutation at G-C pairs in immunoglobulin genes. BMC Immunol *6*, 10. 10.1186/1471-2172-6-10.

85. Grubb, J., Brown, M.S., and Bishop, D.K. (2015). Surface Spreading and Immunostaining of Yeast Chromosomes. J Vis Exp, 53081. 10.3791/53081.

3.7 SUPPLEMENTARY INFORMATION

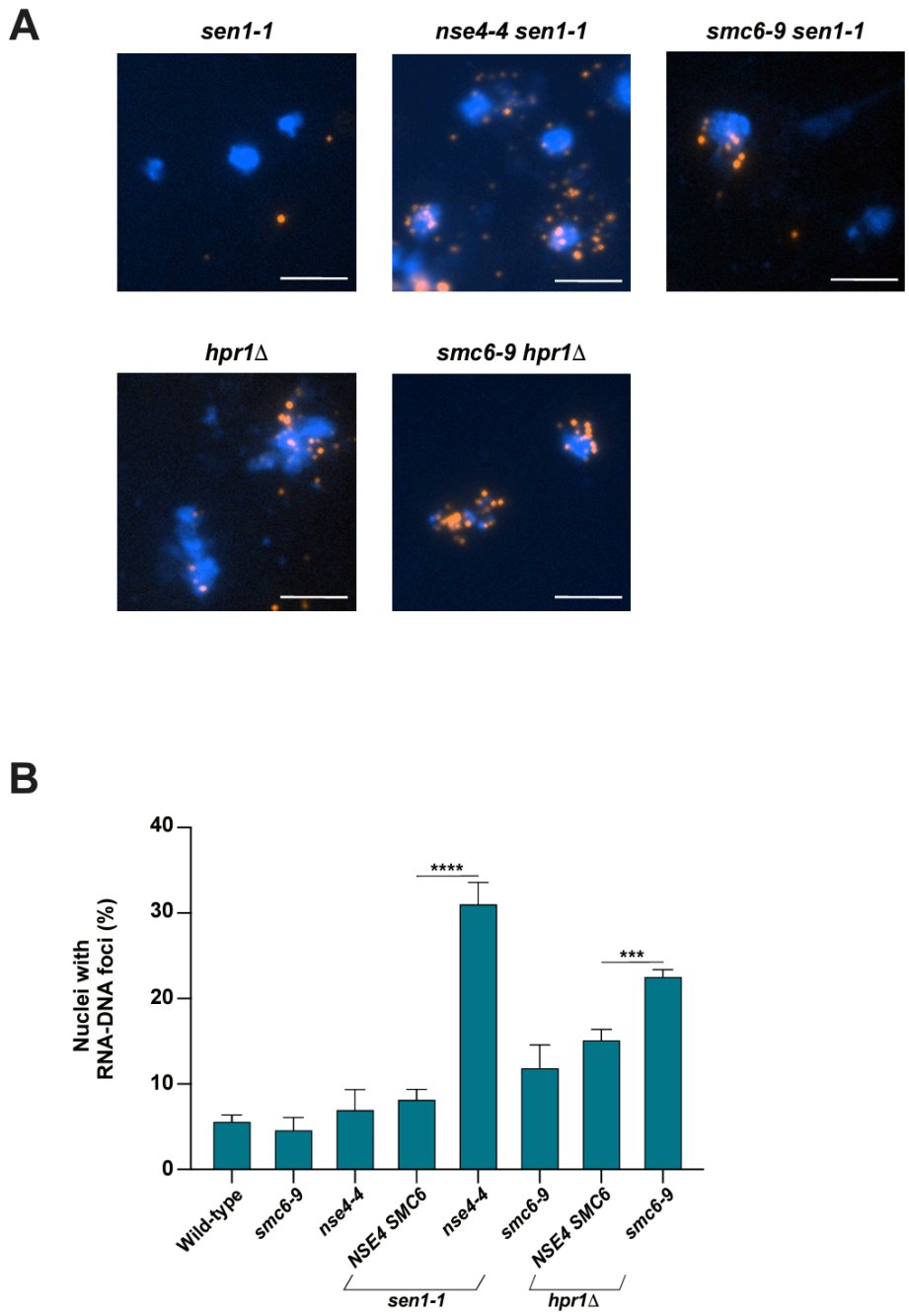


Supplementary Figure 3.S 1: Growth defects of yeast strains carrying mutations in both RNase H and SMC5/6 complex activity. (A) The proliferation capacity of double mutants affecting the Smc5/6 complex and RNase H activity was assessed in the presence of DNA-damaging agents (0.005% MMS and 0.03µM 4NQO at 23°C) (B-C) The proliferation capacity of mutants (as described in Figure 3.1) affecting the Smc5/6 complex and with deletion in only single gene encoding each RNase H enzyme/subunit.



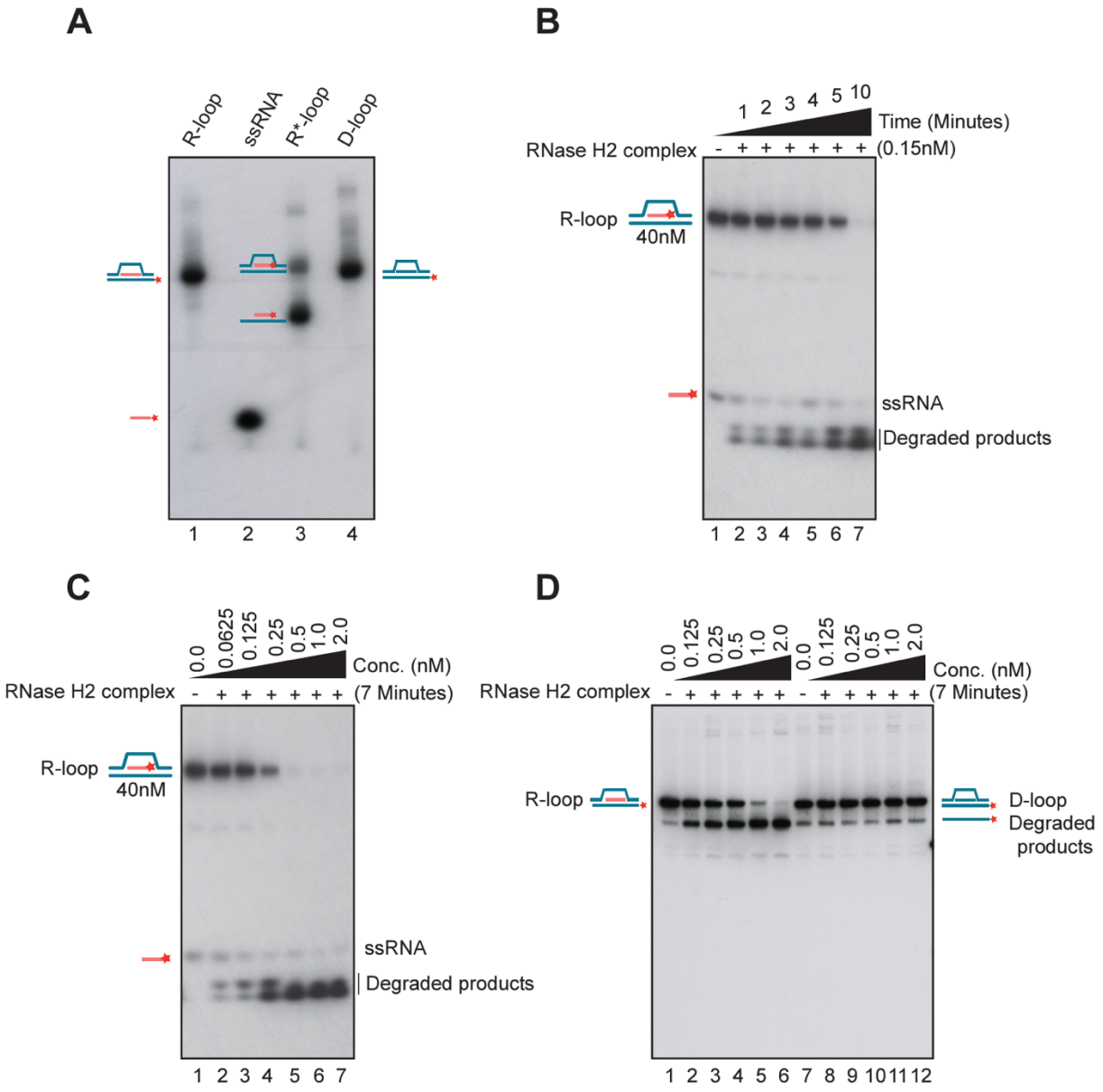
Supplementary Figure 3.S 2: Direct and indirect quantification of R-loops in cells defective for Smc5/6 and RNase H activity. (A) Detection and quantification of RNA-DNA hybrid foci by indirect immunofluorescence microscopy, as described in Figure 3.2. Quantification of nuclei containing (3 to 10 foci) levels of RNA-DNA hybrid structures in single- and double-mutant yeast strains is shown under immunofluorescence micrographs. Scale bar, 5 µm. (B) Detection of R-loops using AID-induced Rad52 foci formation **(extended data from Figure 3.3)**. Representative images of cells carrying Rad52-GFP

foci in the absence (–AID) or presence of AID (+AID) in wild-type and *nse4-4* mutant yeast cells. Scale bar, 10 µm.



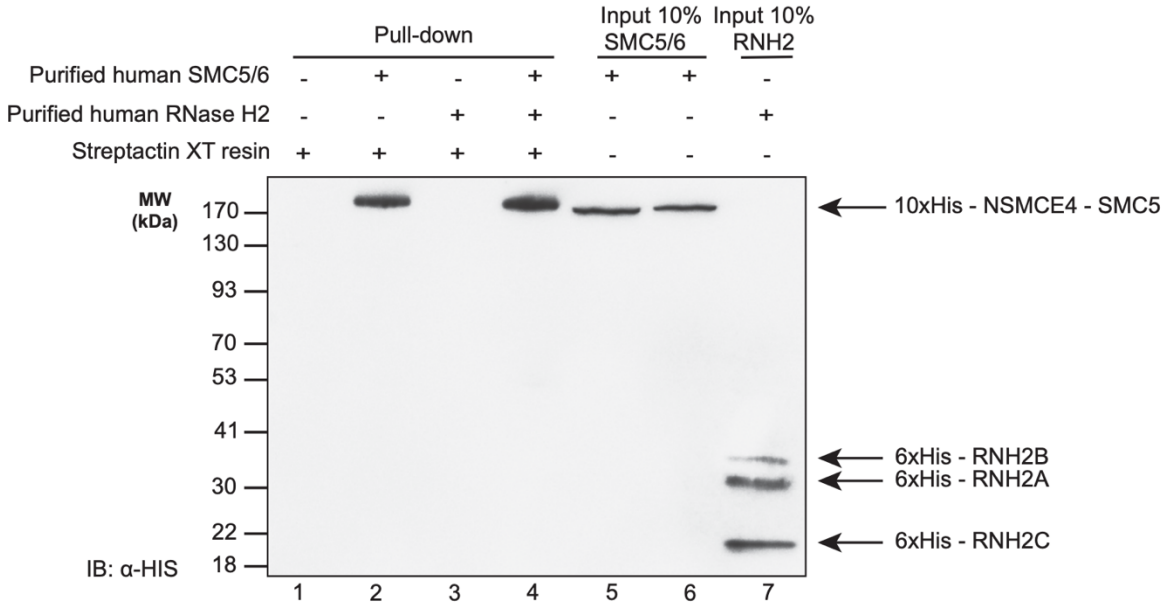
Supplementary Figure 3.S 3: RNA-DNA hybrid accumulation in cells defective for Smc5/6 complex and Sen1 helicase or THO-complex activity. (A-B) The abundance of RNA-DNA hybrids in chromosomes was monitored using the S9.6 antibody by indirect immunofluorescence microscopy on chromosome spreads prepared from WT, single- and double-mutant yeast strains. Representative spreads are shown (A), as shown in Figure 3.2A. Quantification of nuclei containing S9.6 foci is shown (B). At least 100 nuclei were visualized and manually counted for each replicate to obtain the fraction

of nuclei with detectable RNA-DNA hybrids. Data represent mean and SE of three independent experiments. *P* (*) < 0.05; *P* (**) < 0.01; *P* (***) < 0.001 (Student's t-test). Scale bar, 5 µm.



Supplementary Figure 3.S 4: Biochemical properties of nucleic acid substrates and RNase H2 enzyme used in this study. (A) Purification and electrophoretic behavior of ^{32}P-labelled RNA-DNA substrates. Lane 1: R-loop substrate where DNA strand 2 (DD 4264) is radiolabeled at its 5'-end with [γ-^{32}P] ATP; Lane 2: ssRNA oligo (DD 4265) radiolabeled with [γ-^{32}P] ATP; Lane 3: R*-loop substrate (top band) containing a ^{32}P-radiolabelled RNA oligo (DD 4265). Note that a linear DNA-RNA hybrid can be produced as a byproduct of the reannealing reaction (bottom band); Lane 4: D-loop where DNA strand 2 (DD 4264) is radiolabeled at its 5'-end with ^{32}P. All the probes were quantified (nM) using a scintillation counter. The positions of individual probes after migration are marked on both sides of

the gel. (B) An R*-loop substrate containing a ^{32}P-labelled RNA moiety (40 nM) was incubated with 0.15 nM of RNase H2 complex in a time course experiment (t = 1, 2, 3, 4, 5, and 10 minutes). The positions of the substrate and reaction products are marked as described above. Data are representative of 3 independent experiments. (C) The R*-loop substrate described above (40 nM) was incubated with increasing concentrations of RNase H2 complex (0, 0.0625, 0.125, 0.25, 0.5, 1, and 2nM) for 7 min. Data are representative of 3 independent experiments. (D) R-loop degradation assay in the presence of an increasing concentration of RNase H2 complex (as indicated on top of lanes 1-6). Similar reactions were carried out with a D-loop substrate (lanes 7-12) as a negative control for RNase H2 activity. The positions reached by the substrate and reaction products after electrophoresis are marked on both sides of the gels. Data are representative of 3 independent experiments.



Supplementary Figure 3.S 5:**In-vitro binding assay to test the physical interaction between the human SMC5/6 complex and human RNase H2 enzyme.** Representative immunoblot shows the pull-down experiment to test the binding of the SMC5/6 complex and the RNase H2. Lane 1: the negative control with streptactin beads incubated with binding buffer. Lanes 2 and 3: the positive controls with the respective proteins plus beads in the binding buffer (as indicated in lanes 2 and 3). Lane 4: The pull-down shows only the SMC5/6 complex, indicating no detectable interaction between the SMC5/6 complex and the RNase H2. Lane 5-7: 10% input for the Smc5/6 complex (lanes 5 and 6) and the RNase H2 (lanes 7) respectively.

Table 3.S 1: Yeast strains used in this study

Figure	Strain name	Relevant genotype details
Figure 3.1	D7528	*MATa*
	D6531	*MATa rnh1::HIS3MX6 rnh201::kanMX6*
	D6533	*MATa rnh1::HIS3MX6 rnh202::kanMX6*
	D6535	*MATa rnh1::HIS3MX6 rnh203::kanMX6*
	D7799	*MATa nse4-4::URA3*
	D7055	*MATa nse4-4::URA3 rnh1::HIS3MX6 rnh201::kanMX6*
	D7057	*MATa nse4-4::URA3 rnh1::HIS3MX6 rnh202::kanMX6*
	D7084	*MATa nse4-4::URA3 rnh1::HIS3MX6 rnh203::kanMX6*
	D7795	*MATa smc6-9::NAT*
	D7159	*MATa smc6-9::NAT rnh1::HIS3MX6 rnh201::kanMX6*
	D7357	*MATa smc6-9::NAT rnh1::HIS3MX6 rnh202::kanMX6*
	D7157	*MATa smc6-9::NAT rnh1::HIS3MX6 rnh203::kanMX6*
Figure 3.2A	D7528	*MATa*
	D6531	*MATa rnh1::HIS3MX6 rnh201::kanMX6*
	D6533	*MATa rnh1::HIS3MX6 rnh202::kanMX6*
	D6535	*MATa rnh1::HIS3MX6 rnh203::kanMX6*
	D7799	*MATa nse4-4::URA3*
	D7055	*MATa nse4-4::URA3 rnh1::HIS3MX6 rnh201::kanMX6*
	D7057	*MATa nse4-4::URA3 rnh1::HIS3MX6 rnh202::kanMX6*
	D7084	*MATa nse4-4::URA3 rnh1::HIS3MX6 rnh203::kanMX6*
	D7795	*MATa smc6-9::NAT*
	D7159	*MATa smc6-9::NAT rnh1::HIS3MX6 rnh201::kanMX6*
	D7357	*MATa smc6-9::NAT rnh1::HIS3MX6 rnh202::kanMX6*
	D7157	*MATa smc6-9::NAT rnh1::HIS3MX6 rnh203::kanMX6*
Figure 3.2B	D8122	*MATa trp1-1::Pgal1::TRP1*
	D8120	*MATa trp1-1::Pgal1::RNH1::TRP1*
	D8177	*MATa rnh1::HIS3MX6 rnh201::kanMX6 trp1-1::Pgal1::TRP1*
	D8407	*MATa rnh1::HIS3MX6 rnh201::kanMX6 trp1-1::Pgal1::RNH1::TRP1*
	D8128	*MATa smc6-9::NAT rnh1::HIS3MX6 rnh201::kanMX6 trp1-1::Pgal1::TRP1*
	D8119	*MATa smc6-9::NAT rnh1::HIS3MX6 rnh201::kanMX6 trp1-1::Pgal1::*
	D8130	*RNH1::TRP1*
	D8221	*MATa nse4-4::URA3 rnh1::HIS3MX6 rnh203::kanMX6 trp1-1::Pgal1::TRP1*
Figure 3.3A	D7528	*MATa nse4-4::URA3 rnh1::HIS3MX6 rnh203::kanMX6 trp1-1::Pgal1::*
	D6535	*RNH1::TRP1*
	D7084	*MATa*
Figure 3.3B	D8852	*MATa rnh1::HIS3MX6 rnh203::kanMX6*
	D8849	*MATa nse4-4::URA3 rnh1::HIS3MX6 rnh203::kanMX6*
	D8611	*MATa rnh1::HIS3MX6 rnh201::kanMX6 RAD52=EGFP::KanMX6 [p6-YCplac111]*
		MATa rnh1::HIS3MX6 rnh201::kanMX6 nse4-4::URA3 RAD52=EGFP::KanMX6

	D8613	*[p6-YCplac111]*
		MATa rnh1::HIS3MX6 rnh201::kanMX6 RAD52=EGFP::KanMX6 [p1895-pESC-
Figure 3.4	D8457	*LEU-HsAIDSc]*
and 3.S3	D8534	*MATa rnh1::HIS3MX6 rnh201::kanMX6 nse4-4::URA3 RAD52=EGFP::KanMX6*
	D8368	*[p1895-pESC-LEU-HsAIDSc]*
	D8376	*MATa hpr1::HIS3MX6*
	D8438	*MATa smc6-9::NAT hpr1::HIS3MX6*
	D8573	*MATa sen1-1::Tadh1::HIS3MX6*
	D8603	*MATa sen1-1::Tadh1::HIS3MX6 nse4-4::URA3*
	D8601	*MATa sen1-1::Tadh1::HIS3MX6 smc6-9::NAT*
	D6537	*MATa sen1-3[R1605K]::Tadh1::HIS3MX6*
	D8645	*MATa sen1-3[R1605K]::Tadh1::HIS3MX6 nse4-4::URA3*
	D8643	*MATa sen1-3[R1605K]::Tadh1::HIS3MX6 smc6-9::NAT*
	D8691	*MATa pol2[M644G]*
	D8789	*MATa pol2[M644G] nse4-4::URA3*
	D8783	*MATa pol2[M644G] smc6-9::NAT*
	D8844	*MATa rat1-1*
Figure 3.S1A	D7528	*MATa rat1-1 nse4-4::URA3*
	D6531	*MATa rat1-1 smc6-9::NAT*
	D6533	*MATa rat1-1 sen1-1::Tadh1::HIS3MX6 smc6-9::NAT*
	D6535	*MATa*
	D7799	*MATa rnh1::HIS3MX6 rnh201::kanMX6*
	D7055	*MATa rnh1::HIS3MX6 rnh202::kanMX6*
	D7057	*MATa rnh1::HIS3MX6 rnh203::kanMX6*
	D7084	*MATa nse4-4::URA3*
	D7795	*MATa nse4-4::URA3 rnh1::HIS3MX6 rnh201::kanMX6*
	D7159	*MATa nse4-4::URA3 rnh1::HIS3MX6 rnh202::kanMX6*
	D7357	*MATa nse4-4::URA3 rnh1::HIS3MX6 rnh203::kanMX6*
	D7157	*MATa smc6-9::NAT*
Figure 3.S1B	D7528	*MATa smc6-9::NAT rnh1::HIS3MX6 rnh201::kanMX6*
	D9636	*MATa smc6-9::NAT rnh1::HIS3MX6 rnh202::kanMX6*
	D9613	*MATa smc6-9::NAT rnh1::HIS3MX6 rnh203::kanMX6*
	D9607	*MATa*
	D9637	*MATa rnh1::HIS3MX6*
	D9638	*MATa nse4-4::URA3 rnh1::HIS3MX6*
	D9639	*MATa smc6-9::NAT rnh1::HIS3MX6*
	D9615	*MATa rnh201::kanMX6*
	D9617	*MATa rnh202::kanMX6*
	D9620	*MATa rnh203::kanMX6*
	D9605	*MATa nse4-4::URA3 rnh201::kanMX6*
	D9609	*MATa nse4-4::URA3 rnh202::kanMX6*
	D9612	*MATa nse4-4::URA3 rnh203::kanMX6*
Figure 3.S2B	D9105	*MATa smc6-9::NAT rnh201::kanMX6*
	D9101	*MATa smc6-9::NAT rnh202::kanMX6*

	D9107	*MATa smc6-9::NAT rnh203::kanMX6*
	D9105	*MATa RAD52=EGFP::KanMX6 [p6-YCplac111]* *MATa nse4-4::URA3 RAD52=EGFP::KanMX6 [p6-YCplac111]* *MATa RAD52=EGFP::KanMX6 [p1895-pESC-LEU-HsAIDSc]* *MATa nse4-4::URA3 RAD52=EGFP::KanMX6 [p1895-pESC-LEU-HsAIDSc]*

Table 3.S 2: Plasmids used in this study

Figure	Plasmid number	Relevant details
Figure 3.2	**p32**	*YIplac204/GAL1-10*
	p355	*YIplac204_Pgal1_RNH1*
Figure 3.3	**p6**	*YCplac111*
	p1895	*pESC-LEU-HsAIDSc*
Figure 3.6	**p1906**	*pET-hH2ABC*

Chapter IV: Discussion

Maintenance of genome integrity is paramount to proper development and well-being of all living organisms. Cells rely on intricately regulated genomic supervision systems that protect their genomic integrity in the presence of frequent cellular and environmental insults (Groelly *et al*, 2023; Chabanon *et al*, 2021; Zhou & Elledge, 2000). It is becoming increasingly clear that several cellular pathways act simultaneously to safeguard genomic stability and promote error-free cell cycle and division (Shi *et al*, 2022; Bell *et al*, 2011). Despite this progress, how these multiple pathways effectively function together and coordinate their activities in time and space remains a largely unanswered question in the field of genome stability. Importantly, a group of higher-order effectors of chromosome architecture coordinate these multiple pathways towards proper cellular response and regulate the physical organization of the genome to allow distinct molecular transactions.

One class of such higher order effectors include the SMC proteins —a unique group of ATPases that play important and conserved roles in chromosome homeostasis in all domains of life (Kim *et al*, 2023; Bürmann & Löwe, 2023; Bürmann *et al*, 2017). Among them, the Smc5/6 complex plays unique roles in local and context-specific organization of interphase chromosomes with important implications for health and disease (reviewed in Roy *et al*, 2024). While the core biological functions of the other two SMC complexes —cohesin and condensin— are mostly clear, the precise involvement of the Smc5/6 complex in diverse cellular processes is still not completely understood. The findings from this thesis provide novel and essential insights into how the Smc5/6 complex maintains the integrity of genomic DNA in diverse cellular processes.

In this thesis, I first analyzed whether cancer patient mutations found in the SMC5/6 complex can lead to a loss-of-function when inserted at a homologous position in the yeast Smc5/6 complex (Roy *et al*, 2023). Second, I identified a novel function of the Smc5/6 complex in the degradation of toxic R-loop structures —a major concern for genomic instability. Finally, I revealed that the disruption of a subunit interaction hub in Smc6 inactivates the DNA repair activity of the Smc5/6 complex (Appendix). In this chapter, I discuss how the findings of my thesis bridge previous knowledge gaps and identify avenues for future research.

4.1 Investigating the contribution of the SMC5/6 complex in the development of cancer

Until recently, very few studies had explored whether alterations in SMC5/6 complex activity are directly linked to cancer development in humans (Zhou *et al*, 2020; Di Benedetto *et al*, 2022; Saunus *et al*, 2015). In this thesis, I present the first systematic study revealing that the SMC5/6 complex mutations likely contribute to cancer development. My results revealed that the mutations in genes encoding the SMC5/6 complex were present in more than 10% of all cancer patient samples reported in the cBioportal database. Importantly, a majority of patients have alteration in only one of the subunits of the SMC5/6 complex (Chapter II Figure 2.1C). Intriguingly, most of the SMC5/6 complex alterations were heterozygous in nature (Chapter II Figure 2.3A) which further suggests that only one altered allele of the genes encoding SMC5/6 complex is enough to hamper the activity of the complex. This trend was further confirmed while modeling the cancer VUS in yeast model system (Chapter II Figure 2.8). The heterozygous diploid yeast strains carrying wild-type and cancer VUS mutant alleles of Smc5/6 components showed striking growth defects especially when exposed to replication stress. Hence, full functionality of the

Smc5/6 complex requires a tight control over the activity of each subunit of the complex because loss of 50% protein can impart growth defects under heterozygous condition. Overall, these results suggest that mutations in one single subunit particularly in a single allele of the SMC5/6 components can be sufficient to result in detrimental outcomes with potential oncogenic consequences. The exact roles and consequences of SMC5/6 complex mutations/alterations for oncogenesis are not immediately clear due to the contribution of this complex to multiple cellular processes that impact genome stability. Below, I discuss a number of enzymatic behaviours of the mutated SMC5/6 complex that may contribute to genomic instability and subsequent tumorigenesis when inactivated.

Potential roles of NSMCE2 and its SUMO ligase activity in cancer

My findings point towards several putative mechanisms by which altered SMC5/6 complex activity can trigger tumorigenesis. For instance, I showed that NSMCE2 is the most frequently altered subunit in cancer samples with aggressive phenotype —ploidy defects and poor patient survival— of the disease. Other studies also show that *NSMCE2* is highly upregulated in breast tumors with high RNA levels significantly correlating with poor prognosis in cancer patients (Di Benedetto *et al*, 2022). Therefore, this raises the question of how the alteration of the *NSMCE2* gene directly correlates with the ploidy defects in cancer cells and how does that translate to higher mortality in patients. Abnormal ploidy or aneuploidy is a direct consequence of sister-chromatid segregation defects and other forms of mitotic abnormalities (Andriani *et al*, 2017; Potapova & Gorbsky, 2017). Interestingly, the NSMCE2/Mms21 subunit has been previously shown to directly prevent mitotic abnormalities hinting at its possible role in maintaining proper

ploidy in cells (Pond *et al*, 2019; Payne *et al*, 2014). Persistent hydroxyurea treatment in human NSMCE2-deficient cells can lead to collapsed replication forks which might persist into mitosis, leading to increased mitotic DNA damage. It has also been shown that NSMCE2-deficient human cells fail to recruit BLM at stalled replication forks resulting in loss of regulation of RAD51 loading at the replication fork (Pond *et al*, 2019). Subsequently, abnormal RAD51 structure prevents convergence of the active fork with the collapsed fork, resulting in a double fork stall. The resulting double fork stall persists into mitosis where it results in an ultra-fine DNA bridge inhibiting proper chromosome segregation and abnormal ploidy (Pond *et al*, 2019).

Abnormal SUMO ligase activity of the NSMCE2 subunit potentially plays the major role in triggering ploidy defects in cancer patients. Under normal conditions, the SUMO ligase activity of the NSMCE2/Mms21 subunit co-ordinates with the ATPase activity of the holocomplex to promote error-free chromosome disjunction (Bermúdez-López *et al*, 2015). The Mms21-mediated SUMOylation is also essential in resolving HR intermediates during DNA damage repair and replication stress (Solé-Soler & Torres-Rosell, 2020; Roy *et al*, 2024). Unresolved recombination intermediates are major obstacles for the proper separation of sister chromatids before cell division (Baxter, 2015; Kaur *et al*, 2019). Finally, the SUMO-ligase activity of Mms21 is also important for the proper activity of cohesin in mediating sister chromatid cohesion, segregation, and proper disjunction of repetitive sequences in mitosis (Outwin *et al*, 2009; Torres-Rosell *et al*, 2005; Atkins *et al*, 2020; Jacome *et al*, 2015) —an important step towards maintaining proper cellular ploidy. Therefore, it is not surprising that the altered SMC5/6

complex, specifically altered NSMCE2, can result in defective cellular ploidy, leading to tumorigenesis.

Previously it was shown that NSMCE2-mediated SUMOylation is required for G1-S transition in MCF7 breast cancer cells indicating that the manipulation of this activity may alter the growth rates of breast cancer cells (Ni *et al*, 2012). Furthermore, the upregulation of SUMO modification is known to be a prevalent post-translational modification (PTM) in multiple types of cancers hinting towards a significant role of SUMO modification in cancer progression (Zhao *et al*, 2020; Han *et al*, 2018; Lara-Ureña *et al*, 2022; Liu *et al*, 2023b). SUMOylation of proteins involves a complex process fine-tuned by several set of enzymes and proteases to maintain a fine balance between SUMOylation and deSUMOylation. Multiple evidences suggest that SUMOylation imbalance can contribute to several stages of cancer development, including metastasis, angiogenesis, invasion, and proliferation (reviewed in Zhao *et al*, 2020). In light of this, current studies have initiated treatment of tumors by inhibiting the expression of SUMOylation family members, including SUMO E1 or E2 (Huang *et al*, 2024; Du *et al*, 2020; Kroonen & Vertegaal, 2021). Whether the amplification of NSMCE2 observed in my study results in the imbalance of SUMOylation activity and in turn contribute to cancer development and progression, remains one of most important question that needs to be answered in the future.

The SMC5/6 complex and chemotherapeutic drug resistance

Development of drug resistance is a major challenge in cancer chemotherapy. The anthracycline doxorubicin is a known chemotherapeutic drug for many cancers (Christowitz *et al,* 2019;

Ashrafizaveh *et al*, 2021). However, the efficacy of doxorubicin-based regimens is compromised by the frequent occurrence of therapeutic resistance (Christowitz *et al,* 2019; Ashrafizaveh *et al*, 2021). Past genetic-screening of yeast deletion strains identified doxorubicin-resistant mutants which indicated that SUMO pathway is a major determinant of doxorubicin cytotoxicity (Huang *et al*, 2007). Yeast mutants lacking *UBA2* (SUMO activating enzyme; E1), *UBC9* (conjugating enzyme; E2), and *ULP1* and *ULP2* (desumoylation peptidases) are all doxorubicin resistant. Interestingly *mms21* yeast mutants were hypersensitive to doxorubicin unlike the other SUMO-enzyme mutants indicating that tumors with alterations in NSMCE2-SUMO-ligase activity can be effectively treated with doxorubicin mediated chemotherapy (Huang *et al*, 2007).

The ubiquitin-proteasome pathway has also emerged as an important regulator of steady-state protein levels regulating cell cycle, cellular proliferation, apoptosis, and DNA damage response, with major implications in cancer development and drug resistance. Specifically, E3 ubiquitin-ligases are promising drug targets and has been attributed to chemotherapeutic drug resistance in multiple cancers (Anand *et al,* 2023; Narayanan *et al,*2020; Zhuang *et al.* 2019*)*. In that sense, the role of ubiquitin ligase activity of NSMCE1 in chemotherapeutic drug resistance is a potential area of future research. In sum, my research highlights that the molecular knowledge of mutational targets of the SMC5/6 complex can further inform precision therapeutic strategies for cancer patients.

The SMC5/6 complex alterations cause genome stability defects in breast cancer patients

I also show that 18% of breast cancer patients have the SMC5/6 complex alterations (Chapter II Figure 2.1). The predominance of the SMC5/6 complex alterations in breast cancer is intriguing and future studies should focus on the reason behind this specificity towards breast cancer. Germline mutations in a tumor suppressor gene —breast cancer susceptibility gene 1 (BRCA1)— are linked to familial breast and ovarian cancers (Futreal *et al*, 1994; Miki *et al*, 1994). Intriguingly, genetic screening in *Caenorhabditis elegans* suggest that the BRC-1/BRD-1 mediated initiation of HR at stalled replication forks requires SMC5/6 complex activity for resolution of the recombination intermediates (Wolters *et al*, 2014). The SMC5/6 complex also allows the dissociation of the BRC-1/BRD-1 complex from the chromatin (Wolters *et al*, 2014). Hence, the misregulation of the SMC5/6 complex can result in defective unloading of BRC-1/BRD-1 complex, resulting in defective downstream signaling in breast cancer.

The SMC5/6 complex and synthetic partners in cancer

I found that several key genes involved in DNA damage repair, cell cycle and DNA replication were differentially expressed in the SMC5/6 complex altered group of tumor samples (Chapter II Figure 2.5). Hence, the alterations in other chromosome effectors along with the SMC5/6 complex can lead to moderate cellular viability defects contributing to different aggressiveness of cancer subtypes. Furthermore, the SMC5/6 complex alterations can incur sensitivities to several cancer causing genotoxins. A very recent study revealed a synthetic lethal interaction between the loss of SMC5/6 complex activity and activity of APOBEC3A conserved from yeast to humans (O'Leary *et al*, 2023). Cytidine deamination caused by a potent genotoxin APOBEC3 enzymes is among the most common cause of mutagenesis in human cancers (Chan *et al*, 2015). Hence, the SMC5/6

complex presents a potential therapeutic target in tumors with active APOBEC3A (O'Leary *et al*, 2023).

Overall, Chapter II of my thesis provides strong evidence that the genomic instability stemming from altered SMC5/6 complex can result in cancer development. In the future it would be interesting to study the biological and in-depth cellular effects of cancer-mutations of the Smc5/6 complex (identified in my study) at different stages of cancer development and progression.

4.2 The Smc5/6 complex counteracts the harmful effects of R-loop structures which further protects the cell from oncogenic events

The significant oncogenic potential of the SMC5/6 complex observed in my study (Chapter II) may stem from its role in diverse cellular functions such as the repair of DSBs, recovery of stalled replication forks, telomeric length maintenance, and virus restriction. Therefore, the malfunction of the Smc5/6 complex can curb its protective role in genomic maintenance and may generate toxic recombination intermediates *in vivo* in these cellular pathways (Chapter I Figure 1.3). Specifically, most of the nuclear processes outlined above involve the presence of RNA-DNA hybrids or R-loop structures to some extent, which are generally debilitating to overall genomic health. Such observations prompted us to investigate whether RNA based recombination intermediates are prime molecular targets of the Smc5/6 complex in all these pathways (Chapter III). R-loops and its associated transcription–replication conflicts are linked to conditions involving inactivated tumor-suppressor genes or activated oncogenes (Wells *et al*, 2019; Li *et al*, 2023).

Additionally, a complete account of the sequence of molecular events resulting in R-loops removal *in vivo* remains unknown.

In chapter III, I revealed a novel function of the Smc5/6 complex in resolving R-loop and associated genomic instability. I showed that the Smc5/6 complex actively promotes R-loop removal from highly transcribed genes and potentially regulates the function of other R-loop metabolizing factors. Nucleic acid transactions in highly transcribed regions may result in formation of excess RNA-DNA hybrids resulting in replication stress and DNA double stranded breaks. It can be speculated that the Smc5/6 complex may maintain DNA topology at highly transcribed regions in multi-dimensional fashion. First, the Smc5/6 complex work in concert with the R-loop metabolizing factors to stimulate the resolution of RNA-DNA hybrids. Second, it helps in resolving the R-loop associated replication stress and DSBs through homologous recombination. Third, the Smc5/6 complex promotes the resolution of recombination intermediates formed at sites of replication-transcription conflicts (Roy *et al*, 2024). Finally, the Smc5/6 complex associates with transcription-induced positively supercoiled DNA and initiate loop extrusion to maintain the DNA topology for replication restart (Jeppsson *et al*, 2024; Diman *et al*, 2023).

Altogether, the finding of this chapter can be significantly correlated with parallel studies in the field which suggest a crucial role of the Smc5/6 complex in eradicating toxic R-loop structures. Below, I discuss the possible mechanisms by which the Smc5/6 complex prevents the formation of unscheduled R-loops and associated genomic instability.

The Smc5/6 complex is linked to multiple R-loop metabolizing factors

In this thesis, I have shown that the Smc5/6 complex genetically interacts with genes encoding RNase H, Sen1 helicase as well Rat1 which are primary factors responsible for the removal of RNA-DNA hybrid structures (Chapter III). However, the Smc5/6 complex can indirectly contribute to the resolution of RNA-DNA hybrids as well. Loss of Sgs1/BLM activity have also been shown to accumulate R-loops and associated genomic instability across species (Chang *et al*, 2017). In line with this, BLM physically localizes close to DNA-RNA hybrids in human cells and can efficiently unwind R-loops *in vitro* (Chang *et al*, 2017). It is already well-established that Sgs1/BLM is SUMOylated by Mms21 during its recombination function (Bermúdez-López *et al*, 2016). I hypothesize that the Smc5/6 complex- Mms21 mediated SUMOylation of Sgs1/BLM can also play an active role in Sgs1/BLM mediated R-loop resolution. Future research efforts should test this possibility. Additionally, topoisomerases release DNA supercoiling which prevents replication stress at sites of R-loop formation (Promonet *et al.* 2020). The Smc5/6 complex regulates the recruitment of topoisomerases, indirectly participating in prevention of R-loop related replication stress (Jeppsson *et al*, 2014; Bermúdez-López *et al*, 2016). Overall, I speculate that Mms21 subunit can activate several R-loop metabolizing factors through SUMOylation. Moreover, it has also been shown that the Smc5/6 complex regulates the action of Mph1 helicase at sites of RNA-DNA hybrids by inhibiting the Mph1 mediated fork reversal (Chen *et al*, 2009; Lafuente-Barquero *et al*, 2017). To summarize, in the light of recent studies, the Smc5/6 complex can be functionally linked to several R-loop resolving factors. Whether the Smc5/6 complex physically interacts with such factors or modulate their activity by post-translational modifications remains a highly important avenue for future research.

Potential action of the Smc5/6 complex at telomeres

Apart from highly transcribed regions, R-loops are commonly formed at telomeres in the form of long non-coding telomeric RNA transcript TERRA (Rivosecchi *et al*, 2024). I show that the Smc5/6 complex genetically interacts with Rat1 which is primarily involved in RNA-DNA hybrids metabolism at telomeres (Chapter III Figure 3.4). Additionally, Smc5/6 mutants accumulate RNA-DNA hybrid foci at telomeric region without the presence of active RNase H (Chapter III Figure 3.3 and 3.4). In light of this, recently it has been shown that the Flap endonuclease Rad27 also cleaves the RNA moiety of R-loop structures at telomeres preventing telomeric recombination (Liu *et al*, 2023a; Becker *et al*, 2018). Interestingly, it was shown that Rad27 overexpressing yeast cells showed an interesting requirement for the SUMO ligase activity of Mms21 for viability (Becker *et al*, 2018). Hence, the SUMO-ligase activity of Mms21 may promote Rad27 mediated R-loop removal at telomeres.

The R-loop structures can result in multiple forms of genomic instability in cells. The single-stranded DNA of the R-loop structures can often result in the formation of secondary structures like G4 quadruplexes. The Shelterin complex proteins, TRF1 and TRF2 promote the resolution of such G4 quadruplexes at telomeres, and in doing so, prevents the formation of other secondary structures formed as a consequence of R-loop metabolism (Zhang *et al*, 2023). On the other hand, Mms21 SUMOylates multiple telomere-binding proteins, including TRF1 and TRF2 to activate their function (Potts & Yu, 2007). Furthermore, human SLF2-SMC5 stimulate proper efficient replication through G4-quadruplex DNA secondary structures putatively at sites of RNA-DNA hybrids formed everywhere in the genome (Grange *et al*, 2022). Hence, the Smc5/6 complex not

only promotes resolution of replication stress and DSBs but also can prevent the formation of secondary structures formed at sites of R-loops. Taken together, these observations indicate that the Smc5/6 complex can achieve the resolution of R-loops at telomeres by diverse approaches.

R-loop resolution by the Smc5/6 complex at ribosomal DNA

I have also revealed that simultaneous loss of Smc5/6 complex and RNase H activity are associated with R-loop accumulation at the ribosomal DNA (rDNA) locus (Chapter III Figure 3.3). RNA polymerase I is the primary polymerase responsible for rDNA transcription and consequently, misregulation of this enzyme can result in elevated transcription of rDNA repeats, ultimately resulting in RNA-DNA hybrid accumulation (Goodfellow *et al*, 2013). Interestingly, Nse1-mediated ubiquitination is associated with ribosome biogenesis and metabolism (Ibars *et al*, 2023). Specifically, Nse1-E3 ligase activity ubiquitinylate RNA polymerase I/RPA190 and subsequently degrade RNA polymerase I (Ibars *et al*, 2023). Hence, Rpa190 ubiquitination by Nse1 can regulate the activity of RNA polymerase I and may thus prevent RNA polymerase I mediated R-loop formation at this locus.

I also observed elevation of R-loop foci in mitochondrial DNA regions in Smc5/6-RNaseH double mutants. In this context, loss of BRCA2 —an important regulator of HR— induces oxidative stress which further impairs RNaseH1 recruitment to mitochondrial DNA (Renaudin *et al*, 2021). This results in R-loop formation and stalled mitochondrial DNA replication. It can be speculated that the Smc5/6 complex can also regulate BRCA2 during homologous recombination for efficient removal of R-loops at mitochondrial DNA.

Finally, SUMOylation plays an important role in mRNA metabolism including transcription and post-transcriptional processing (e.g., capping, splicing and polyadenylation), failure of which can trigger unscheduled R-loop formation (Cao *et al*, 2023). Hence, Mms21 mediated SUMOylation have potential function in R-loop resolution. In summary, Chapter III provides a detailed study of how the Smc5/6 complex can activate and regulate multiple pathways to remove toxic R-loop structures, a common source of endogenous toxicity in cells. Future studies should also focus on investigating the mode of regulation induced by the Smc5/6 complex at R-loop sites. Additionally, identification of the structural conformation of the Smc5/6 holocomplex that supports the loading of the complex to R-loop structures and whether that the mode of binding regulates the downstream effects of R-loop degradation activity are important avenues of future work. In this vein, the Nse5/6 sub-complex —the chromatin loaders of the Smc5/6 complex— might also play a vital role in R-loop homeostasis (Li *et al*, 2023). Moreover, it would be interesting to explore if any differences exist in the sequence of events and mechanism by which the Smc5/6 complex promotes R-loop repair under both normal and stressful conditions. Future investigations should elucidate the protective roles of the Smc5/6 complex against the harmful consequence of R-loop structures.

4.3. Identification of a critical hub within the subunits of the Smc5/6 complex (Appendix)

The incomplete understanding of the molecular mechanism of the Smc5/6 complex over the years was partially due to the unsuccessful attempts in isolating pure SMC5/6 complex to perform biochemical and structural studies. A previous study from our laboratory (Appendix of

this thesis) was the first to report the development of a unique co-translational folding strategy to purify the human SMC5/6 complex in its active form (Serrano *et al*, 2020). This was achieved in parallel to another study that successfully purified the octameric yeast Smc5/6 holocomplex (Gutierrez-Escribano *et al*, 2020). Following the purification of the active human SMC5/6 complex, we were able to perform complex biochemical and structural analyses to understand the biological activity of the SMC5/6 complex.

We hypothesized that the physiological activity of the multi-subunit Smc5/6 complex depends on crucial interaction interfaces connecting different subunits within the complex. I focussed on understanding how such interactions affect its overall functionality during DNA repair. I identified a novel interaction hub within Smc6 that connects most of the subunits of the complex and regulates Smc5/6 complex activity and its structural topology. The results indicated a possible destabilization of Smc6 ATPase head/neck region that can affect the structural integrity of the complex. The clinical and biological implications of our discoveries can be directly linked to the development of the chromosome breakage syndromes (e.g., LIC syndrome) in humans (van der Crabben *et al*, 2016) because the mutations in LIC syndrome patients mapped directly to the region of human NSE3 that associates with this interaction hub in SMC6 (van der Crabben *et al*, 2016; Serrano *et al*, 2020).

Our results focussed on the interaction of the subunits in the pentameric complex (SMC5-6-NSE1-3-4). The structural regulation of the Smc5/6 complex depends majorly on the regulatory activity of the SUMO-ligase Nse2 and the chromatin loaders Nse5/6 subcomplex. In line with this, the

latest cryo-EM structure of the yeast octameric Smc5/6 complex evidently indicates that the chromatin loaders of the Smc5/6 complex, the Nse5/6 subcomplex interacts with the ATPase head domain and the adjacent coiled-coil neck region of Smc6 —putative SMC6 interaction hub in our study— (Chapter I, Figure 1.1C). Intriguingly, Nse6 competes with Nse4 in its binding to the neck region of Smc6 and thus regulates the ATPase activity depending on the biological requirement of the complex (Li *et al*, 2023; Taschner & Gruber, 2023; Yu *et al*, 2022). However, the binding of Nse4 to the Smc6 functions as a stable linker to support the engagement of the ATPase head domain. Genetic data suggest that defective Nse6-Smc6 neck interaction affects HJ resolution in yeast in presence of inactivated enzyme Mms4. On the other hand, defective Nse6-Smc6 does not affect growth of cells lacking the replication termination regulator Rrm3 indicating that the Nse6-Smc6 neck interaction is specifically crucial for HJ resolution (Li *et al*, 2023). Whether the Smc6/Nse6 interaction hub can also regulate the Nse5/6 mediated viral DNA restriction needs to be elucidated. Overall, I infer from these observations that patients with LIC syndrome might harbor a defect in a NSE /SMC6 gate that controls the Smc5/6 complex configuration and transit of DNA into the Smc5/6 ring structure.

4.4 Conclusion

The work presented here has wide-ranging implications in connecting the fundamental biology of one of the most enigmatic Structural Maintenance of Chromosomes complexes —the Smc5/6 complex— to its roles in human disease conditions. My thesis not only provides novel insights on the core molecular functions of the Smc5/6 complex during maintenance of genomic stability but also reveals how altered Smc5/6 complex activity can lead to tumorigenesis and other human

conditions involving chromosomal abnormalities. Taken together, my work emphasizes the importance of basic research in promoting the mission of translational medicine and improving the likelihood of identifying novel and effective clinical treatments. My discoveries also reveal that focusing research on high-level chromosomal effectors, such as the Smc5/6 complex, can aid in the development of innovative therapeutic strategies. Together, the body of the work presented in this thesis makes an impactful contribution to the field of genome stability.

References

Abdul F, Diman A, Baechler B, Ramakrishnan D, Kornyeyev D, Beran RK, Fletcher SP & Strubin M (2022a) Smc5/6 silences episomal transcription by a three-step function. *Nat Struct Mol Biol* 29: 922–931

Adamus M, Lelkes E, Potesil D, Ganji SR, Kolesar P, Zabrady K, Zdrahal Z & Palecek JJ (2020) Molecular Insights into the Architecture of the Human SMC5/6 Complex. *J Mol Biol* 432: 3820–3837

Agashe S, Joseph CR, Reyes TAC, Menolfi D, Giannattasio M, Waizenegger A, Szakal B & Branzei D (2021) Smc5/6 functions with Sgs1-Top3-Rmi1 to complete chromosome replication at natural pause sites. *Nat Commun* 12: 2111

Albuquerque CP, Wang G, Lee NS, Kolodner RD, Putnam CD & Zhou H (2013) Distinct SUMO Ligases Cooperate with Esc2 and Slx5 to Suppress Duplication-Mediated Genome Rearrangements. *PLOS Genetics* 9: e1003670

Alt A, Dang HQ, Wells OS, Polo LM, Smith MA, McGregor GA, Welte T, Lehmann AR, Pearl LH, Murray JM, *et al* (2017) Specialized interfaces of Smc5/6 control hinge stability and DNA association. *Nature Communications* 8: 14011

Ampatzidou E, Irmisch A, O'Connell MJ & Murray JM (2006) Smc5/6 Is Required for Repair at Collapsed Replication Forks. *Molecular and Cellular Biology* 26: 9387–9401

Anand, S., Nedeva, C., Chitti, S.V. et al. (2023) The E3 ubiquitin ligase NEDD4 regulates chemoresistance to 5-fluorouracil in colorectal cancer cells by altering JNK signalling. Cell Death Dis 14, 828.

Anderson DE, Losada A, Erickson HP & Hirano T (2002) Condensin and cohesin display different arm conformations with characteristic hinge angles. *J Cell Biol* 156: 419–424

Andrews EA, Palecek J, Sergeant J, Taylor E, Lehmann AR & Watts FZ (2005) Nse2, a Component of the Smc5-6 Complex, Is a SUMO Ligase Required for the Response to DNA Damage. *Molecular and Cellular Biology* 25: 185–196

Andriani GA, Vijg J & Montagna C (2017) Mechanisms and consequences of aneuploidy and chromosome instability in the aging brain. *Mechanisms of Ageing and Development* 161: 19–36

Appanah R, Lones EC, Aiello U, Libri D & De Piccoli G (2020) Sen1 Is Recruited to Replication Forks via Ctf4 and Mrc1 and Promotes Genome Stability. *Cell Rep* 30: 2094-2105.e9

Arnould C, Rocher V, Finoux A-L, Clouaire T, Li K, Zhou F, Caron P, Mangeot PhilippeE, Ricci EP, Mourad R, *et al* (2021) Loop extrusion as a mechanism for DNA damage repair foci formation. *Nature* 590: 660–665

Ashrafizaveh S, Ashrafizadeh M, Zarrabi A, Husmandi K, Zabolian A, Shahinozzaman M, et al. (2021)Long non-coding RNA in the doxorubicin resistance of cancer cells. Cancer Lett.

Atkins A, Xu MJ, Li M, Rogers NP, Pryzhkova MV & Jordan PW (2020a) SMC5/6 is required for replication fork stability and faithful chromosome segregation during neurogenesis. *Elife* 9: e61171

Barth R, Pradhan B, Kim E, Davidson IF, van der Torre J, Peters J-M & Dekker C (2023) Testing pseudotopological and nontopological models for SMC-driven DNA loop extrusion against roadblock-traversal experiments. *Sci Rep* 13: 8100

Baxter J (2015) "Breaking Up Is Hard to Do": The Formation and Resolution of Sister Chromatid Intertwines. *Journal of Molecular Biology* 427: 590–607

Becker JR, Gallo D, Leung W, Croissant T, Thu YM, Nguyen HD, Starr TK, Brown GW & Bielinsky A-K (2018) Flap endonuclease overexpression drives genome instability and DNA damage hypersensitivity in a PCNA-dependent manner. *Nucleic Acids Res* 46: 5634–5650

Behlke-Steinert S, Touat-Todeschini L, Skoufias DA & Margolis RL (2009) SMC5 and MMS21 are required for chromosome cohesion and mitotic progression. *Cell Cycle* 8: 2211–2218

Bell O, Tiwari VK, Thomä NH & Schübeler D (2011) Determinants and dynamics of genome accessibility. *Nat Rev Genet* 12: 554–564

Bentley P, Tan MJA, McBride AA, White EA & Howley PM (2018) The SMC5/6 Complex Interacts with the Papillomavirus E2 Protein and Influences Maintenance of Viral Episomal DNA. *J Virol* 92: e00356-18

Bermejo R, Doksani Y, Capra T, Katou Y-M, Tanaka H, Shirahige K & Foiani M (2007) 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

Bermudez-Lopez M, Ceschia A, de Piccoli G, Colomina N, Pasero P, Aragon L & Torres-Rosell J (2010) The Smc5/6 complex is required for dissolution of DNA-mediated sister chromatid linkages. *Nucleic Acids Research* 38: 6502–6512

Bermúdez-López M, Pociño-Merino I, Sánchez H, Bueno A, Guasch C, Almedawar S, Bru-Virgili S, Garí E, Wyman C, Reverter D, *et al* (2015) ATPase-Dependent Control of the Mms21 SUMO Ligase during DNA Repair. *PLOS Biology* 13: e1002089

Bermúdez-López M, Villoria MT, Esteras M, Jarmuz A, Torres-Rosell J, Clemente-Blanco A & Aragon L (2016) Sgs1's roles in DNA end resection, HJ dissolution, and crossover suppression require a two-step SUMO regulation dependent on Smc5/6. *Genes Dev* 30: 1339–1356

Betts Lindroos H, Ström L, Itoh T, Katou Y, Shirahige K & Sjögren C (2006) Chromosomal Association of the Smc5/6 Complex Reveals that It Functions in Differently Regulated Pathways. *Molecular Cell* 22: 755–767

Bonner JN, Choi K, Xue X, Torres NP, Szakal B, Wei L, Wan B, Arter M, Matos J, Sung P, *et al* (2016) Smc5/6 Mediated Sumoylation of the Sgs1-Top3-Rmi1 Complex Promotes Removal of Recombination Intermediates. *Cell Rep* 16: 368–378

Branzei D, Vanoli F & Foiani M (2008) SUMOylation regulates Rad18-mediated template switch. *Nature* 456: 915–920

Bürmann F, Basfeld A, Vazquez Nunez R, Diebold-Durand M-L, Wilhelm L & Gruber S (2017) Tuned SMC Arms Drive Chromosomal Loading of Prokaryotic Condensin. *Mol Cell* 65: 861-872.e9

Bürmann F, Lee B-G, Than T, Sinn L, O'Reilly FJ, Yatskevich S, Rappsilber J, Hu B, Nasmyth K & Löwe J (2019) A folded conformation of MukBEF and cohesin. *Nat Struct Mol Biol* 26: 227–236

Bürmann F & Löwe J (2023) Structural biology of SMC complexes across the tree of life. *Curr Opin Struct Biol* 80: 102598

Bustard DE, Menolfi D, Jeppsson K, Ball LG, Dewey SC, Shirahige K, Sjögren C, Branzei D & Cobb JA (2012) During Replication Stress, Non-Smc Element 5 (Nse5) Is Required for Smc5/6 Protein Complex Functionality at Stalled Forks. *J Biol Chem* 287: 11374–11383

Cao Y, Huang C, Zhao X & Yu J (2023) Regulation of SUMOylation on RNA metabolism in cancers. *Front Mol Biosci* 10: 1137215

Cardozo Gizzi AM, Cattoni DI, Fiche J-B, Espinola SM, Gurgo J, Messina O, Houbron C, Ogiyama Y, Papadopoulos GL, Cavalli G, *et al* (2019) Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Mol Cell* 74: 212-222.e5

Caridi CP, D'Agostino C, Ryu T, Zapotoczny G, Delabaere L, Li X, Khodaverdian VY, Amaral N, Lin E, Rau AR, *et al* (2018) Nuclear F-actin and myosins drive relocalization of heterochromatic breaks. *Nature* 559: 54–60

Carusillo A & Mussolino C (2020) DNA Damage: From Threat to Treatment. *Cells* 9: 1665

Chabanon RM, Rouanne M, Lord CJ, Soria J-C, Pasero P & Postel-Vinay S (2021) Targeting the DNA damage response in immuno-oncology: developments and opportunities. *Nat Rev Cancer* 21: 701–717

Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, Kim J, Kwiatkowski DJ, Fargo DC, Mieczkowski PA, *et al* (2015) An APOBEC3A hypermutation signature is distinguishable

from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat Genet* 47: 1067–1072

Chang EY-C, Novoa CA, Aristizabal MJ, Coulombe Y, Segovia R, Chaturvedi R, Shen Y, Keong C, Tam AS, Jones SJM, *et al* (2017) RECQ-like helicases Sgs1 and BLM regulate R-loop–associated genome instability. *J Cell Biol* 216: 3991–4005

Chang JT-H, Li S, Beckwitt EC, Than T, Haluska C, Chandanani J, O'Donnell ME, Zhao X & Liu S (2022) Smc5/6's multifaceted DNA binding capacities stabilize branched DNA structures. *Nat Commun* 13: 7179

Chen D, Dundr M, Wang C, Leung A, Lamond A, Misteli T & Huang S (2004) Condensed mitotic chromatin is accessible to transcription factors and chromatin structural proteins. *Journal of Cell Biology* 168: 41–54

Chen Y-H, Choi K, Szakal B, Arenz J, Duan X, Ye H, Branzei D & Zhao X (2009) Interplay between the Smc5/6 complex and the Mph1 helicase in recombinational repair. *PNAS* 106: 21252–21257

Cheutin T, McNairn AJ, Jenuwein T, Gilbert DM, Singh PB & Misteli T (2003) Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science* 299: 721–725

Chiolo I, Minoda A, Colmenares SU, Polyzos A, Costes SV & Karpen GH (2011) Double-Strand Breaks in Heterochromatin Move Outside of a Dynamic HP1a Domain to Complete Recombinational Repair. *Cell* 144: 732–744

Choi K, Szakal B, Chen Y-H, Branzei D & Zhao X (2010) The Smc5/6 Complex and Esc2 Influence Multiple Replication-associated Recombination Processes in *Saccharomyces cerevisiae*. *MBoC* 21: 2306–2314

Christowitz, C., Davis, T., Isaacs, A. et al (2019). Mechanisms of doxorubicin-induced drug resistance and drug resistant tumour growth in a murine breast tumour model. BMC Cancer 19, 757.

Cost GJ & Cozzarelli NR (2006) Smc5p Promotes Faithful Chromosome Transmission and DNA Repair in *Saccharomyces cerevisiae*. *Genetics* 172: 2185–2200

van der Crabben SN, Hennus MP, McGregor GA, Ritter DI, Nagamani SCS, Wells OS, Harakalova M, Chinn IK, Alt A, Vondrova L, *et al* (2016) Destabilized SMC5/6 complex leads to chromosome breakage syndrome with severe lung disease. *Journal of Clinical Investigation* 126: 2881–2892

Davidson IF, Bauer B, Goetz D, Tang W, Wutz G & Peters J-M (2019) DNA loop extrusion by human cohesin. *Science* 366: 1338–1345

Davidson IF & Peters J-M (2021) Genome folding through loop extrusion by SMC complexes. *Nat Rev Mol Cell Biol* 22: 445–464

De Bont R & van Larebeke N (2004) Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* 19: 169–185

Decorsière A, Mueller H, van Breugel PC, Abdul F, Gerossier L, Beran RK, Livingston CM, Niu C, Fletcher SP, Hantz O, *et al* (2016) Hepatitis B virus X protein identifies the Smc5/6 complex as a host restriction factor. *Nature* 531: 386–389

Deep A, Gu Y, Gao Y-Q, Ego KM, Herzik MA, Zhou H & Corbett KD (2022) The SMC-family Wadjet complex protects bacteria from plasmid transformation by recognition and cleavage of closed-circular DNA. *Molecular Cell* 82: 4145-4159.e7

Degtyareva NP, Heyburn L, Sterling J, Resnick MA, Gordenin DA & Doetsch PW (2013) Oxidative stress-induced mutagenesis in single-strand DNA occurs primarily at cytosines and is DNA polymerase zeta-dependent only for adenines and guanines. *Nucleic Acids Res* 41: 8995–9005

Dekker J & Mirny L (2016) The 3D Genome as Moderator of Chromosomal Communication. *Cell* 164: 1110–1121

Di Benedetto C, Oh J, Choudhery Z, Shi W, Valdes G & Betancur P (2022) NSMCE2, a novel super-enhancer-regulated gene, is linked to poor prognosis and therapy resistance in breast cancer. *BMC Cancer* 22: 1056

Diman A, Panis G, Castrogiovanni C, Prados J, Baechler B & Strubin M (2023) Human Smc5/6 recognises transcription-generated positive DNA supercoils. 2023.05.04.539344 doi:10.1101/2023.05.04.539344 [PREPRINT]

Doyle JM, Gao J, Wang J, Yang M & Potts PR (2010) MAGE-RING protein complexes comprise a family of E3 ubiquitin ligases. *Mol Cell* 39: 963–974

Du L, Fakih MG, Rosen ST & Chen Y (2020) SUMOylation of E2F1 Regulates Expression of EZH2. *Cancer Research* 80: 4212–4223

Duan X, Yang Y, Chen Y-H, Arenz J, Rangi GK, Zhao X & Ye H (2009) Architecture of the Smc5/6 Complex of Saccharomyces cerevisiae Reveals a Unique Interaction between the Nse5-6 Subcomplex and the Hinge Regions of Smc5 and Smc6. *J Biol Chem* 284: 8507–8515

Duina AA, Miller ME & Keeney JB (2014) Budding Yeast for Budding Geneticists: A Primer on the Saccharomyces cerevisiae Model System. *Genetics* 197: 33–48

Dupont L, Bloor S, Williamson JC, Cuesta SM, Shah R, Teixeira-Silva A, Naamati A, Greenwood EJD, Sarafianos SG, Matheson NJ, *et al* (2021) The SMC5/6 complex compacts and silences unintegrated HIV-1 DNA and is antagonized by Vpr. *Cell Host Microbe* 29: 792-805.e6

Dvořák Tomaštíková E, Prochazkova K, Yang F, Jemelkova J, Finke A, Dorn A, Said M, Puchta H & Pecinka A (2023) SMC5/6 complex-mediated SUMOylation stimulates DNA-protein cross-link repair in Arabidopsis. *Plant Cell* 35: 1532–1547

Emerson DJ, Zhao PA, Cook AL, Barnett RJ, Klein KN, Saulebekova D, Ge C, Zhou L, Simandi Z, Minsk MK, *et al* (2022) Cohesin-mediated loop anchors confine the locations of human replication origins. *Nature* 606: 812–819

Etheridge TJ, Villahermosa D, Campillo-Funollet E, Herbert AD, Irmisch A, Watson AT, Dang HQ, Osborne MA, Oliver AW, Carr AM, *et al* (2021) Live-cell single-molecule tracking highlights requirements for stable Smc5/6 chromatin association in vivo. *Elife* 10: e68579

Fujioka Y, Kimata Y, Nomaguchi K, Watanabe K & Kohno K (2002) Identification of a novel non-structural maintenance of chromosomes (SMC) component of the SMC5-SMC6 complex involved in DNA repair. *J Biol Chem* 277: 21585–21591

Funato K, Otsuka M, Sekiba K, Miyakawa Y, Seimiya T, Shibata C, Kishikawa T & Fujishiro M (2022) Hepatitis B virus-associated hepatocellular carcinoma with Smc5/6 complex deficiency is susceptible to PARP inhibitors. *Biochem Biophys Res Commun* 607: 89–95

Futreal PA, Liu Q, Shattuck-Eidens D, Cochran C, Harshman K, Tavtigian S, Bennett LM, Haugen-Strano A, Swensen J & Miki Y (1994) BRCA1 mutations in primary breast and ovarian carcinomas. *Science* 266: 120–122

Gallego-Paez LM, Tanaka H, Bando M, Takahashi M, Nozaki N, Nakato R, Shirahige K & Hirota T (2014) Smc5/6-mediated regulation of replication progression contributes to chromosome assembly during mitosis in human cells. *Mol Biol Cell* 25: 302–317

Ganji M, Shaltiel IA, Bisht S, Kim E, Kalichava A, Haering CH & Dekker C (2018) Real-time imaging of DNA loop extrusion by condensin. *Science* 360: 102–105

García Fernández F & Fabre E (2022) The Dynamic Behavior of Chromatin in Response to DNA Double-Strand Breaks. *Genes (Basel)* 13: 215

Gibson RT & Androphy EJ (2020) The SMC5/6 Complex Represses the Replicative Program of High-Risk Human Papillomavirus Type 31. *Pathogens* 9: 786

Golloshi R, Sanders JT & McCord RP (2017) Genome organization during the cell cycle: unity in division. *Wiley Interdiscip Rev Syst Biol Med* 9

Gomez R, Jordan PW, Viera A, Alsheimer M, Fukuda T, Jessberger R, Llano E, Pendas AM, Handel MA & Suja JA (2013) Dynamic localization of SMC5/6 complex proteins during mammalian meiosis and mitosis suggests functions in distinct chromosome processes. *Journal of Cell Science* 126: 4239–4252

Goodfellow SJ, Zomerdijk JC (2013) Basic mechanisms in RNA polymerase I transcription of the ribosomal RNA genes. *Subcell Biochem*. ;61:211-36.

Grange LJ, Reynolds JJ, Ullah F, Isidor B, Shearer RF, Latypova X, Baxley RM, Oliver AW, Ganesh A, Cooke SL, *et al* (2022) Pathogenic variants in SLF2 and SMC5 cause segmented chromosomes and mosaic variegated hyperploidy. *Nat Commun* 13: 6664

Groelly FJ, Fawkes M, Dagg RA, Blackford AN & Tarsounas M (2023) Targeting DNA damage response pathways in cancer. *Nat Rev Cancer* 23: 78–94

Guacci V, Koshland D & Strunnikov A (1997) A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in S. cerevisiae. *Cell* 91: 47–57

Gutierrez-Escribano P, Hormeño S, Madariaga-Marcos J, Solé-Soler R, O'Reilly FJ, Morris K, Aicart-Ramos C, Aramayo R, Montoya A, Kramer H, *et al* (2020a) Purified Smc5/6 Complex Exhibits DNA Substrate Recognition and Compaction. *Mol Cell* 80: 1039-1054.e6

Complex Exhibits DNA Substrate Recognition and Compaction. *Mol Cell* 80: 1039-1054.e6

Haering CH, Löwe J, Hochwagen A & Nasmyth K (2002) Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol Cell* 9: 773–788

Hahm JY, Park J, Jang E-S & Chi SW (2022) 8-Oxoguanine: from oxidative damage to epigenetic and epitranscriptional modification. *Exp Mol Med* 54: 1626–1642

Hakem R (2008) DNA-damage repair; the good, the bad, and the ugly. *EMBO J* 27: 589–605

Hallett ST, Campbell Harry I, Schellenberger P, Zhou L, Cronin NB, Baxter J, Etheridge TJ, Murray JM & Oliver AW (2022) Cryo-EM structure of the Smc5/6 holo-complex. *Nucleic Acids Res* 50: 9505–9520

Hallett ST, Schellenberger P, Zhou L, Beuron F, Morris E, Murray JM & Oliver AW (2021) Nse5/6 is a negative regulator of the ATPase activity of the Smc5/6 complex. *Nucleic Acids Res* 49: 4534–4549

Han C, Zhang D, Gui C, Huang L, Chang S, Dong L, Bai L, Wu S & Lan K (2022) KSHV RTA antagonizes SMC5/6 complex-induced viral chromatin compaction by hijacking the ubiquitin-proteasome system. *PLoS Pathog* 18: e1010744

Han Z-J, Feng Y-H, Gu B-H, Li Y-M & Chen H (2018) The post-translational modification, SUMOylation, and cancer (Review). *Int J Oncol* 52: 1081–1094

Hang LE, Peng J, Tan W, Szakal B, Menolfi D, Sheng Z, Lobachev K, Branzei D, Feng W & Zhao X (2015) Rtt107 Is a Multi-functional Scaffold Supporting Replication Progression with Partner SUMO and Ubiquitin Ligases. *Mol Cell* 60: 268–279

Hansen AS, Cattoglio C, Darzacq X & Tjian R (2018) Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* 9: 20–32

Hartwell LH (1967) Macromolecule synthesis in temperature-sensitive mutants of yeast. *J Bacteriol* 93: 1662–1670

Hauer MH & Gasser SM (2017) Chromatin and nucleosome dynamics in DNA damage and repair. *Genes Dev* 31: 2204–2221

Hays M, Young JM, Levan PF & Malik HS (2020) A natural variant of the essential host gene MMS21 restricts the parasitic 2-micron plasmid in Saccharomyces cerevisiae. *Elife* 9: e62337

Hazbun TR, Malmström L, Anderson S, Graczyk BJ, Fox B, Riffle M, Sundin BA, Aranda JD, McDonald WH, Chiu C-H, *et al* (2003) Assigning Function to Yeast Proteins by Integration of Technologies. *Molecular Cell* 12: 1353–1365

Heideker J, Prudden J, Perry JJP, Tainer JA & Boddy MN (2011) SUMO-Targeted Ubiquitin Ligase, Rad60, and Nse2 SUMO Ligase Suppress Spontaneous Top1–Mediated DNA Damage and Genome Instability. *PLoS Genet* 7: e1001320

Helleday T, Eshtad S & Nik-Zainal S (2014) Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* 15: 585–598

Herskowitz I (1988) Life cycle of the budding yeast Saccharomyces cerevisiae. *Microbiol Rev* 52: 536–553

Higashi TL & Uhlmann F (2022) SMC complexes: Lifting the lid on loop extrusion. *Curr Opin Cell Biol* 74: 13–22

Hirano T, Kobayashi R & Hirano M (1997) Condensins, chromosome condensation protein complexes containing XCAP-C, XCAP-E and a Xenopus homolog of the Drosophila Barren protein. *Cell* 89: 511–521

Hoencamp C & Rowland BD (2023) Genome control by SMC complexes. *Nat Rev Mol Cell Biol*

Horigome C, Bustard DE, Marcomini I, Delgoshaie N, Tsai-Pflugfelder M, Cobb JA & Gasser SM (2016) PolySUMOylation by Siz2 and Mms21 triggers relocation of DNA breaks to nuclear pores through the Slx5/Slx8 STUbL. *Genes Dev* 30: 931–945

Horváth A, Rona G, Pagano M & Jordan PW (2020) Interaction between NSMCE4A and GPS1 links the SMC5/6 complex to the COP9 signalosome. *BMC Molecular and Cell Biology* 21: 36

Hu B, Liao C, Millson SH, Mollapour M, Prodromou C, Pearl LH, Piper PW & Panaretou B (2005) Qri2/Nse4, a component of the essential Smc5/6 DNA repair complex. *Molecular Microbiology* 55: 1735–1750

Huang C-H, Yang T-T & Lin K-I (2024) Mechanisms and functions of SUMOylation in health and disease: a review focusing on immune cells. *Journal of Biomedical Science* 31: 16

Huang R-X & Zhou P-K (2020) DNA damage response signaling pathways and targets for radiotherapy sensitization in cancer. *Sig Transduct Target Ther* 5: 1–27

Huang R-Y, Kowalski D, Minderman H, Gandhi N & Johnson ES (2007) Small ubiquitin-related modifier pathway is a major determinant of doxorubicin cytotoxicity in Saccharomyces cerevisiae. *Cancer Res* 67: 765–772

Ibars E, Codina-Fabra J, Bellí G, Casas C, Tarrés M, Solé-Soler R, Lorite NP, Ximénez-Embún P, Muñoz J, Colomina N, *et al* (2023) Ubiquitin proteomics identifies RNA polymerase I as a target of the Smc5/6 complex. *Cell Reports* 42: 112463

Irmisch A, Ampatzidou E, Mizuno K, O'Connell MJ & Murray JM (2009) Smc5/6 maintains stalled replication forks in a recombination-competent conformation. *EMBO J* 28: 144–155

Irwan ID, Bogerd HP & Cullen BR (2022) Epigenetic silencing by the SMC5/6 complex mediates HIV-1 latency. *Nat Microbiol* 7: 2101–2113

Jacome A, Gutierrez-Martinez P, Schiavoni F, Tenaglia E, Martinez P, Rodríguez-Acebes S, Lecona E, Murga M, Méndez J, Blasco MA, *et al* (2015) NSMCE2 suppresses cancer and aging in mice independently of its SUMO ligase activity. *The EMBO Journal* 34: 2604–2619

Jeggo PA, Pearl LH & Carr AM (2016) DNA repair, genome stability and cancer: a historical perspective. *Nat Rev Cancer* 16: 35–42

Jeppsson K, Carlborg KK, Nakato R, Berta DG, Lilienthal I, Kanno T, Lindqvist A, Brink MC, Dantuma NP, Katou Y, *et al* (2014) The Chromosomal Association of the Smc5/6 Complex Depends on Cohesion and Predicts the Level of Sister Chromatid Entanglement. *PLoS Genet* 10: e1004680

Jeppsson K, Pradhan B, Sutani T, Sakata T, Umeda Igarashi M, Berta DG, Kanno T, Nakato R, Shirahige K, Kim E, Björkegren C. (2024) Loop-extruding Smc5/6 organizes transcription-induced positive DNA supercoils. Mol Cell. Mar 7;84(5):867-882.e5.

Jo A, Li S, Shin JW, Zhao X & Cho Y (2021) Structure Basis for Shaping the Nse4 Protein by the Nse1 and Nse3 Dimer within the Smc5/6 Complex. *J Mol Biol* 433: 166910

Kanno T, Berta DG & Sjögren C (2015) The Smc5/6 Complex Is an ATP-Dependent Intermolecular DNA Linker. *Cell Reports* 12: 1471–1482

Kaur H, GN K & Lichten M (2019) Unresolved Recombination Intermediates Cause a RAD9-Dependent Cell Cycle Arrest in Saccharomyces cerevisiae. *Genetics* 213: 805–818

Kegel A, Betts-Lindroos H, Kanno T, Jeppsson K, Ström L, Katou Y, Itoh T, Shirahige K & Sjögren C (2011) Chromosome length influences replication-induced topological stress. *Nature* 471: 392–396

Kegel A & Sjögren C (2010) The Smc5/6 Complex: More Than Repair? *Cold Spring Harb Symp Quant Biol* 75: 179–187

Kim E, Barth R & Dekker C (2023) Looping the Genome with SMC Complexes. *Annu Rev Biochem* 92: 15–41

Kolesar P, Stejskal K, Potesil D, Murray JM & Palecek JJ (2022) Role of Nse1 Subunit of SMC5/6 Complex as a Ubiquitin Ligase. *Cells* 11: 165

Kong M, Cutts EE, Pan D, Beuron F, Kaliyappan T, Xue C, Morris EP, Musacchio A, Vannini A & Greene EC (2020) Human Condensin I and II Drive Extensive ATP-Dependent Compaction of Nucleosome-Bound DNA. *Mol Cell* 79: 99-114.e9

Kozakova L, Vondrova L, Stejskal K, Charalabous P, Kolesar P, Lehmann AR, Uldrijan S, Sanderson CM, Zdrahal Z & Palecek JJ (2015) The melanoma-associated antigen 1 (MAGEA1) protein stimulates the E3 ubiquitin-ligase activity of TRIM31 within a TRIM31-MAGEA1-NSE4 complex. *Cell Cycle* 14: 920–930

Kroonen JS & Vertegaal ACO (2021) Targeting SUMO Signaling to Wrestle Cancer. *Trends in Cancer* 7: 496–510

Laflamme G, Tremblay-Boudreault T, Roy M-A, Andersen P, Bonneil É, Atchia K, Thibault P, D'Amours D & Kwok BH (2014) Structural maintenance of chromosome (SMC) proteins link microtubule stability to genome integrity. *J Biol Chem* 289: 27418–27431

Lafuente-Barquero J, Luke-Glaser S, Graf M, Silva S, Gómez-González B, Lockhart A, Lisby M, Aguilera A & Luke B (2017) The Smc5/6 complex regulates the yeast Mph1 helicase at RNA-DNA hybrid-mediated DNA damage. *PLOS Genetics* 13: e1007136

Lara-Ureña N, Jafari V & García-Domínguez M (2022) Cancer-Associated Dysregulation of Sumo Regulators: Proteases and Ligases. *Int J Mol Sci* 23: 8012

Lee B-G, Merkel F, Allegretti M, Hassler M, Cawood C, Lecomte L, O'Reilly FJ, Sinn LR, Gutierrez-Escribano P, Kschonsak M, *et al* (2020) Cryo-EM structures of holo condensin reveal a subunit flip-flop mechanism. *Nat Struct Mol Biol* 27: 743–751

Lehmann AR, Walicka M, Griffiths DJ, Murray JM, Watts FZ, McCready S & Carr AM (1995) The rad18 gene of Schizosaccharomyces pombe defines a new subgroup of the SMC superfamily involved in DNA repair. *Mol Cell Biol* 15: 7067–7080

Leung GP, Lee L, Schmidt TI, Shirahige K & Kobor MS (2011) Rtt107 is required for recruitment of the SMC5/6 complex to DNA double strand breaks. *J Biol Chem* 286: 26250–26257

Li F, Zafar A, Luo L, Denning AM, Gu J, Bennett A, Yuan F & Zhang Y (2023a) R-Loops in Genome Instability and Cancer. *Cancers (Basel)* 15: 4986

Li S, Mutchler A, Zhu X, So S, Epps J, Guan D, Zhao X & Xue X (2022) Multifaceted regulation of the sumoylation of the Sgs1 DNA helicase. *J Biol Chem* 298: 102092

Li S, Yu Y, Zheng J, Miller-Browne V, Ser Z, Kuang H, Patel DJ & Zhao X (2023) Molecular basis for Nse5-6 mediated regulation of Smc5/6 functions. *Proceedings of the National Academy of Sciences* 120: e2310924120

Lindahl T & Barnes DE (2000) Repair of Endogenous DNA Damage. *Cold Spring Harb Symp Quant Biol* 65: 127–134

Liu C-C, Chan H-R, Su G-C, Hsieh Y-Z, Lei K-H, Kato T, Yu T-Y, Kao Y-W, Cheng T-H, Chi P, *et al* (2023a) Flap endonuclease Rad27 cleaves the RNA of R-loop structures to suppress telomere recombination. *Nucleic Acids Res* 51: 4398–4414

Liu Y, Li Y & Lu X (2016) Regulators in the DNA damage response. *Archives of Biochemistry and Biophysics* 594: 18–25

Liu Y, Liu J, Peng N, Hai S, Zhang S, Zhao H & Liu W (2023b) Role of non-canonical post-translational modifications in gastrointestinal tumors. *Cancer Cell International* 23: 225

Martin RM & Cardoso MC (2010) Chromatin condensation modulates access and binding of nuclear proteins. *FASEB J* 24: 1066–1072

McAleenan A, Cordon-Preciado V, Clemente-Blanco A, Liu I-C, Sen N, Leonard J, Jarmuz A & Aragón L (2012) SUMOylation of the α-kleisin subunit of cohesin is required for DNA damage-induced cohesion. *Curr Biol* 22: 1564–1575

McDonald WH, Pavlova Y, Yates JR & Boddy MN (2003) Novel Essential DNA Repair Proteins Nse1 and Nse2 Are Subunits of the Fission Yeast Smc5-Smc6 Complex. *J Biol Chem* 278: 45460–45467

Meng X, Wei L, Peng XP & Zhao X (2019) Sumoylation of the DNA polymerase ε by the Smc5/6 complex contributes to DNA replication. *PLOS Genetics* 15: e1008426

Menolfi D, Delamarre A, Lengronne A, Pasero P & Branzei D (2015) Essential Roles of the Smc5/6 Complex in Replication through Natural Pausing Sites and Endogenous DNA Damage Tolerance. *Molecular Cell* 60: 835–846

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

Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM & Ding W (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266: 66–71

Misteli T & Soutoglou E (2009) The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nat Rev Mol Cell Biol* 10: 243–254

Moon K-W & Ryu J-K (2023) Current working models of SMC-driven DNA-loop extrusion. *Biochem Soc Trans* 51: 1801–1810

Moradi-Fard S, Mojumdar A, Chan M, Harkness TAA & Cobb JA (2021) Smc5/6 in the rDNA modulates lifespan independently of Fob1. *Aging Cell* 20: e13373

Moradi-Fard S, Sarthi J, Tittel-Elmer M, Lalonde M, Cusanelli E, Chartrand P & Cobb JA (2016) Smc5/6 Is a Telomere-Associated Complex that Regulates Sir4 Binding and TPE. *PLOS Genetics* 12: e1006268

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

Murray JM & Carr AM (2008) Smc5/6: a link between DNA repair and unidirectional replication? *Nat Rev Mol Cell Biol* 9: 177–182

Narayanan S, Cai CY, Assaraf YG, Guo HQ, Cui Q, Wei L, Huang JJ, Ashby CR Jr, Chen ZS. (2020) Targeting the ubiquitin-proteasome pathway to overcome anti-cancer drug resistance. Drug Resist Updat. Jan;48:100663.

Nakamura K, Saredi G, Becker JR, Foster BM, Nguyen NV, Beyer TE, Cesa LC, Faull PA, Lukauskas S, Frimurer T, *et al* (2019) H4K20me0 recognition by BRCA1–BARD1 directs homologous recombination to sister chromatids. *Nat Cell Biol* 21: 311–318

Nasim A & Smith BP (1975) Genetic control of radiation sensitivity in Schizosaccharomyces pombe. *Genetics* 79: 573–582

Ni H-J, Chang Y-N, Kao P-H, Chai S-P, Hsieh Y-H, Wang D-H & Fong JC (2012) Depletion of SUMO ligase hMMS21 impairs G1 to S transition in MCF-7 breast cancer cells. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1820: 1893–1900

Nie H, Wang Y, Yang X, Liao Z, He X, Zhou J & Ou C (2021) Clinical Significance and Integrative Analysis of the SMC Family in Hepatocellular Carcinoma. *Front Med (Lausanne)* 8: 727965

Noël J-F & Wellinger RJ (2011) Abrupt telomere losses and reduced end-resection can explain accelerated senescence of Smc5/6 mutants lacking telomerase. *DNA Repair (Amst)* 10: 271–282

Norbury CJ & Hickson ID (2001) Cellular Responses to DNA Damage. *Annual Review of Pharmacology and Toxicology* 41: 367–401

O'Leary DR, Hansen AR, Fingerman DF, Tran T, Harris BR, Hayer KE, Tennekoon M, DeWeerd RA, Meroni A, Szeto JH, *et al* (2023) The SMC5/6 complex is required for maintenance of genome integrity upon APOBEC3A-mediated replication stress. 2023.11.28.568952 doi:10.1101/2023.11.28.568952 [PREPRINT]

Oravcová M & Boddy MN (2019) Recruitment, loading, and activation of the Smc5-Smc6 SUMO ligase. *Curr Genet* 65: 669–676

Oravcová M, Gadaleta MC, Nie M, Reubens MC, Limbo O, Russell P & Boddy MN (2019) Brc1 Promotes the Focal Accumulation and SUMO Ligase Activity of Smc5-Smc6 during Replication Stress. *Mol Cell Biol* 39: e00271-18

Oravcová M, Nie M, Zilio N, Maeda S, Jami-Alahmadi Y, Lazzerini-Denchi E, Wohlschlegel JA, Ulrich HD, Otomo T & Boddy MN (2022) The Nse5/6-like SIMC1-SLF2 complex localizes SMC5/6 to viral replication centers. *Elife* 11: e79676

Oshidari R, Strecker J, Chung DKC, Abraham KJ, Chan JNY, Damaren CJ & Mekhail K (2018) Nuclear microtubule filaments mediate non-linear directional motion of chromatin and promote DNA repair. *Nat Commun* 9: 2567

Outwin EA, Irmisch A, Murray JM & O'Connell MJ (2009) Smc5-Smc6-Dependent Removal of Cohesin from Mitotic Chromosomes. *Mol Cell Biol* 29: 4363–4375

Palecek J, Vidot S, Feng M, Doherty AJ & Lehmann AR (2006) The Smc5-Smc6 DNA Repair Complex BRIDGING OF THE Smc5-Smc6 HEADS BY THE KLEISIN, Nse4, AND NON-KLEISIN SUBUNITS. *J Biol Chem* 281: 36952–36959

Palecek JJ (2018) SMC5/6: Multifunctional Player in Replication. *Genes (Basel)* 10: 7

Payne F, Colnaghi R, Rocha N, Seth A, Harris J, Carpenter G, Bottomley WE, Wheeler E, Wong S, Saudek V, *et al* (2014) Hypomorphism in human NSMCE2 linked to primordial dwarfism and insulin resistance. *J Clin Invest* 124: 4028–4038

Pebernard S, McDonald WH, Pavlova Y, Yates JR & Boddy MN (2004) Nse1, Nse2, and a Novel Subunit of the Smc5-Smc6 Complex, Nse3, Play a Crucial Role in Meiosis. *MBoC* 15: 4866–4876

Pebernard S, Schaffer L, Campbell D, Head SR & Boddy MN (2008) Localization of Smc5/6 to centromeres and telomeres requires heterochromatin and SUMO, respectively. *EMBO J* 27: 3011–3023

Pebernard S, Wohlschlegel J, McDonald WH, Yates JR & Boddy MN (2006) The Nse5-Nse6 Dimer Mediates DNA Repair Roles of the Smc5-Smc6 Complex. *Molecular and Cellular Biology* 26: 1617–1630

Pedroza-Garcia JA, Xiang Y & De Veylder L (2022) Cell cycle checkpoint control in response to DNA damage by environmental stresses. *The Plant Journal* 109: 490–507

Peng XP, Lim S, Li S, Marjavaara L, Chabes A & Zhao X (2018) Acute Smc5/6 depletion reveals its primary role in rDNA replication by restraining recombination at fork pausing sites. *PLoS Genet* 14: e1007129

Peng XP & Zhao X (2023) The multi-functional Smc5/6 complex in genome protection and disease. *Nat Struct Mol Biol* 30: 724–734

Perera D, Poulos RC, Shah A, Beck D, Pimanda JE & Wong JWH (2016) Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes. *Nature* 532: 259–263

Petermann E, Lan L & Zou L (2022) Sources, resolution and physiological relevance of R-loops and RNA-DNA hybrids. *Nat Rev Mol Cell Biol* 23: 521–540

Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH & Ferrin TE (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* 30: 70–82

Pond KW, Renty C de, Yagle MK & Ellis NA (2019) Rescue of collapsed replication forks is dependent on NSMCE2 to prevent mitotic DNA damage. *PLOS Genetics* 15: e1007942

Postow L, Crisona NJ, Peter BJ, Hardy CD & Cozzarelli NR (2001) Topological challenges to DNA replication: conformations at the fork. *Proc Natl Acad Sci U S A* 98: 8219–8226

Potapova T & Gorbsky GJ (2017) The Consequences of Chromosome Segregation Errors in Mitosis and Meiosis. *Biology (Basel)* 6: 12

Potenski CJ & Klein HL (2014) How the misincorporation of ribonucleotides into genomic DNA can be both harmful and helpful to cells. *Nucleic Acids Res* 42: 10226–10234

Potts PR, Porteus MH & Yu H (2006) Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks. *EMBO J* 25: 3377–3388

Potts PR & Yu H (2005) Human MMS21/NSE2 is a SUMO ligase required for DNA repair. *Mol Cell Biol* 25: 7021–7032

Potts PR & Yu H (2007) The SMC5/6 complex maintains telomere length in ALT cancer cells through SUMOylation of telomere-binding proteins. *Nature Structural &Amp; Molecular Biology* 14: 581

Pradhan B, Kanno T, Umeda Igarashi M, Loke MS, Baaske MD, Wong JSK, Jeppsson K, Björkegren C & Kim E (2023) The Smc5/6 complex is a DNA loop-extruding motor. *Nature* 616: 843–848

Prakash L & Prakash S (1977) Isolation and characterization of MMS-sensitive mutants of Saccharomyces cerevisiae. *Genetics* 86: 33–55

Promonet, A., Padioleau, I., Liu, Y. et al. (2020) Topoisomerase 1 prevents replication stress at R-loop-enriched transcription termination sites. Nat Commun 11, 3940.

Räschle M, Smeenk G, Hansen RK, Temu T, Oka Y, Hein MY, Nagaraj N, Long DT, Walter JC, Hofmann K, *et al* (2015) DNA repair. Proteomics reveals dynamic assembly of repair complexes during bypass of DNA cross-links. *Science* 348: 1253671

Renaudin X, Lee M, Shehata M, Surmann E-M & Venkitaraman AR (2021) BRCA2 deficiency reveals that oxidative stress impairs RNaseH1 function to cripple mitochondrial DNA maintenance. *Cell Rep* 36: 109478

Rivosecchi J, Jurikova K & Cusanelli E (2024) Telomere-specific regulation of TERRA and its impact on telomere stability. *Seminars in Cell & Developmental Biology* 157: 3–23

Romero-Laorden N & Castro E (2017) Inherited mutations in DNA repair genes and cancer risk. *Curr Probl Cancer* 41: 251–264

Rossi F, Helbling-Leclerc A, Kawasumi R, Jegadesan NK, Xu X, Devulder P, Abe T, Takata M, Xu D, Rosselli F, *et al* (2020) SMC5/6 acts jointly with Fanconi anemia factors to support DNA repair and genome stability. *EMBO reports* 21: e48222

Roy M-A & D'Amours D (2011) DNA-binding properties of Smc6, a core component of the Smc5-6 DNA repair complex. *Biochem Biophys Res Commun* 416: 80–85

Roy M-A, Dhanaraman T & D'Amours D (2015) The Smc5-Smc6 heterodimer associates with DNA through several independent binding domains. *Scientific Reports* 5: 9797

Roy M-A, Siddiqui N & D'Amours D (2011) Dynamic and selective DNA-binding activity of Smc5, a core component of the Smc5-Smc6 complex. *Cell Cycle* 10: 690–700

Roy S, Adhikary H & D'Amours D (2024) The SMC5/6 complex: folding chromosomes back into shape when genomes take a break. *Nucleic Acids Research*: gkae103

Roy S, Zaker A, Mer A & D'Amours D (2023) Large-scale phenogenomic analysis of human cancers uncovers frequent alterations affecting SMC5/6 complex components in breast cancer. *NAR Cancer* 5: zcad047

Ryu T, Spatola B, Delabaere L, Bowlin K, Hopp H, Kunitake R, Karpen GH & Chiolo I (2015) Heterochromatic breaks move to the nuclear periphery to continue recombinational repair. *Nat Cell Biol* 17: 1401–1411

Saunus JM, Quinn MCJ, Patch A-M, Pearson JV, Bailey PJ, Nones K, McCart Reed AE, Miller D, Wilson PJ, Al-Ejeh F, *et al* (2015) Integrated genomic and transcriptomic analysis of human brain metastases identifies alterations of potential clinical significance. *J Pathol* 237: 363–378

Schumacher B, Pothof J, Vijg J & Hoeijmakers JHJ (2021) The central role of DNA damage in the ageing process. *Nature* 592: 695–703

Sekiba K, Otsuka M, Funato K, Miyakawa Y, Tanaka E, Seimiya T, Yamagami M, Tsutsumi T, Okushin K, Miyakawa K, *et al* (2022) HBx-induced degradation of Smc5/6 complex impairs homologous recombination-mediated repair of damaged DNA. *Journal of Hepatology* 76: 53–62

Sergeant J, Taylor E, Palecek J, Fousteri M, Andrews EA, Sweeney S, Shinagawa H, Watts FZ & Lehmann AR (2005) Composition and Architecture of the Schizosaccharomyces pombe Rad18 (Smc5-6) Complex. *Molecular and Cellular Biology* 25: 172–184

Serrano D, Cordero G, Kawamura R, Sverzhinsky A, Sarker M, Roy S, Malo C, Pascal JM, Marko JF & D'Amours D (2020) The Smc5/6 Core Complex Is a Structure-Specific DNA Binding and Compacting Machine. *Mol Cell* 80: 1025-1038.e5

Shaltiel IA, Datta S, Lecomte L, Hassler M, Kschonsak M, Bravo S, Stober C, Ormanns J, Eustermann S & Haering CH (2022) A hold-and-feed mechanism drives directional DNA loop extrusion by condensin. *Science* 376: 1087–1094

Sharma R, Lewis S & Wlodarski MW (2020) DNA Repair Syndromes and Cancer: Insights Into Genetics and Phenotype Patterns. *Front Pediatr* 8

Sherwood K, Ward JC, Soriano I, Martin L, Campbell A, Rahbari R, Kafetzopoulos I, Sproul D, Green A, Sampson JR, *et al* (2023) Germline de novo mutations in families with Mendelian cancer syndromes caused by defects in DNA repair. *Nat Commun* 14: 3636

Shi X, Zhai Z, Chen Y, Li J & Nordenskiöld L (2022) Recent Advances in Investigating Functional Dynamics of Chromatin. *Front Genet* 13: 870640

Shintomi K, Inoue F, Watanabe H, Ohsumi K, Ohsugi M & Hirano T (2017) Mitotic chromosome assembly despite nucleosome depletion in Xenopus egg extracts. *Science* 356: 1284–1287

Smith MJ, Bryant EE, Joseph FJ & Rothstein R (2019) DNA damage triggers increased mobility of chromosomes in G1-phase cells. *Mol Biol Cell* 30: 2620–2625

Smith MJ, Bryant EE & Rothstein R (2018) Increased chromosomal mobility after DNA damage is controlled by interactions between the recombination machinery and the checkpoint. *Genes Dev* 32: 1242–1251

Soh Y-M, Bürmann F, Shin H-C, Oda T, Jin KS, Toseland CP, Kim C, Lee H, Kim SJ, Kong M-S, *et al* (2015) Molecular basis for SMC rod formation and its dissolution upon DNA binding. *Mol Cell* 57: 290–303

Solé-Soler R & Torres-Rosell J (2020) Smc5/6, an atypical SMC complex with two RING-type subunits. *Biochem Soc Trans* 48: 2159–2171

Sollier J, Driscoll R, Castellucci F, Foiani M, Jackson SP & Branzei D (2009) The Saccharomyces cerevisiae Esc2 and Smc5-6 Proteins Promote Sister Chromatid Junction-mediated Intra-S Repair. *MBoC* 20: 1671–1682

Ström L & Sjögren C (2007) Chromosome segregation and double-strand break repair — a complex connection. *Current Opinion in Cell Biology* 19: 344–349

Tanasie N-L, Gutiérrez-Escribano P, Jaklin S, Aragon L & Stigler J (2022) Stabilization of DNA fork junctions by Smc5/6 complexes revealed by single-molecule imaging. *Cell Rep* 41: 111778

Taschner M, Basquin J, Steigenberger B, Schäfer IB, Soh Y-M, Basquin C, Lorentzen E, Räschle M, Scheltema RA & Gruber S (2021) Nse5/6 inhibits the Smc5/6 ATPase and modulates DNA substrate binding. *EMBO J* 40: e107807

Taschner M & Gruber S (2023) DNA segment capture by Smc5/6 holocomplexes. *Nat Struct Mol Biol* 30: 619–628

Taylor EM, Copsey AC, Hudson JJR, Vidot S & Lehmann AR (2008) Identification of the Proteins, Including MAGEG1, That Make Up the Human SMC5-6 Protein Complex. *Molecular and Cellular Biology* 28: 1197–1206

Thattikota Y, Tollis S, Palou R, Vinet J, Tyers M & D'Amours D (2018) Cdc48/VCP Promotes Chromosome Morphogenesis by Releasing Condensin from Self-Entrapment in Chromatin. *Mol Cell* 69: 664-676.e5

Tian H, Gao Z, Li H, Zhang B, Wang G, Zhang Q, Pei D & Zheng J (2015) DNA damage response--a double-edged sword in cancer prevention and cancer therapy. *Cancer Lett* 358: 8–16

Torres-Rosell J, Machín F, Farmer S, Jarmuz A, Eydmann T, Dalgaard JZ & Aragón L (2005) SMC5 and SMC6 genes are required for the segregation of repetitive chromosome regions. *Nat Cell Biol* 7: 412–419

Torres-Rosell J, Sunjevaric I, De Piccoli G, Sacher M, Eckert-Boulet N, Reid R, Jentsch S, Rothstein R, Aragón L & Lisby M (2007) The Smc5-Smc6 complex and SUMO modification of Rad52 regulates recombinational repair at the ribosomal gene locus. *Nat Cell Biol* 9: 923–931

Tubbs A & Nussenzweig A (2017) Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* 168: 644–656

Varejão N, Ibars E, Lascorz J, Colomina N, Torres-Rosell J & Reverter D (2018) DNA activates the Nse2/Mms21 SUMO E3 ligase in the Smc5/6 complex. *EMBO J* 37

Venegas AB, Natsume T, Kanemaki M & Hickson ID (2020) Inducible Degradation of the Human SMC5/6 Complex Reveals an Essential Role Only during Interphase. *Cell Reports* 31: 107533

Verver DE, Zheng Y, Speijer D, Hoebe R, Dekker HL, Repping S, Stap J & Hamer G (2016) Non-SMC Element 2 (NSMCE2) of the SMC5/6 Complex Helps to Resolve Topological Stress. *International Journal of Molecular Sciences* 17: 1782

Vondrova L, Kolesar P, Adamus M, Nociar M, Oliver AW & Palecek JJ (2020) A role of the Nse4 kleisin and Nse1/Nse3 KITE subunits in the ATPase cycle of SMC5/6. *Sci Rep* 10: 9694

Wang JC (2002) Cellular roles of DNA topoisomerases: a molecular perspective. *Nature Reviews Molecular Cell Biology* 3: 430

Wells JP, White J & Stirling PC (2019) R Loops and Their Composite Cancer Connections. *Trends Cancer* 5: 619–631

Weon JL, Yang SW & Potts PR (2018) Cytosolic Iron-Sulfur Assembly Is Evolutionarily Tuned by a Cancer-Amplified Ubiquitin Ligase. *Mol Cell* 69: 113-125.e6

Whalen JM, Dhingra N, Wei L, Zhao X & Freudenreich CH (2020) Relocation of Collapsed Forks to the Nuclear Pore Complex Depends on Sumoylation of DNA Repair Proteins and Permits Rad51 Association. *Cell Rep* 31: 107635

Winczura A, Appanah R, Tatham MH, Hay RT & Piccoli GD (2019) The S phase checkpoint promotes the Smc5/6 complex dependent SUMOylation of Pol2, the catalytic subunit of DNA polymerase ε. *PLOS Genetics* 15: e1008427

Wolters S, Ermolaeva MA, Bickel JS, Fingerhut JM, Khanikar J, Chan RC & Schumacher B (2014) Loss of Caenorhabditis elegans BRCA1 Promotes Genome Stability During Replication in smc-5 Mutants. *Genetics* 196: 985–999

Wu N, Kong X, Ji Z, Zeng W, Potts PR, Yokomori K & Yu H (2012) Scc1 sumoylation by Mms21 promotes sister chromatid recombination through counteracting Wapl. *Genes Dev* 26: 1473–1485

Wutz G, Várnai C, Nagasaka K, Cisneros DA, Stocsits RR, Tang W, Schoenfelder S, Jessberger G, Muhar M, Hossain MJ, *et al* (2017) Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J* 36: 3573–3599

Xu P, Sun D, Gao Y, Jiang Y, Zhong M, Zhao G, Chen J, Wang Z, Liu Q, Hong J, *et al* (2021) Germline mutations in a DNA repair pathway are associated with familial colorectal cancer. *JCI Insight* 6: e148931

Xu W, Ma C, Zhang Q, Zhao R, Hu D, Zhang X, Chen J, Liu F, Wu K, Liu Y, *et al* (2018) PJA1 Coordinates with the SMC5/6 Complex To Restrict DNA Viruses and Episomal Genes in an Interferon-Independent Manner. *J Virol* 92: e00825-18

Yang H, Gao S, Chen J & Lou W (2020) UBE2I promotes metastasis and correlates with poor prognosis in hepatocellular carcinoma. *Cancer Cell Int* 20: 234

Yiu SPT, Guo R, Zerbe C, Weekes MP & Gewurz BE (2022) Epstein-Barr virus BNRF1 destabilizes SMC5/6 cohesin complexes to evade its restriction of replication compartments. *Cell Rep* 38: 110411

Yong-Gonzales V, Hang LE, Castellucci F, Branzei D & Zhao X (2012) The Smc5-Smc6 Complex Regulates Recombination at Centromeric Regions and Affects Kinetochore Protein Sumoylation during Normal Growth. *PLOS ONE* 7: e51540

Yu Y, Li S, Ser Z, Kuang H, Than T, Guan D, Zhao X & Patel DJ (2022) Cryo-EM structure of DNA-bound Smc5/6 reveals DNA clamping enabled by multi-subunit conformational changes. *Proc Natl Acad Sci U S A* 119: e2202799119

Yu Y, Li S, Ser Z, Sanyal T, Choi K, Wan B, Kuang H, Sali A, Kentsis A, Patel DJ, *et al* (2021) Integrative analysis reveals unique structural and functional features of the Smc5/6 complex. *Proc Natl Acad Sci U S A* 118: e2026844118

Yuen KC & Gerton JL (2018) Taking cohesin and condensin in context. *PLOS Genetics* 14: e1007118

Zabrady K, Adamus M, Vondrova L, Liao C, Skoupilova H, Novakova M, Jurcisinova L, Alt A, Oliver AW, Lehmann AR, *et al* (2016) Chromatin association of the SMC5/6 complex is dependent on binding of its NSE3 subunit to DNA. *Nucleic Acids Res* 44: 1064–1079

Zapatka M, Pociño-Merino I, Heluani-Gahete H, Bermúdez-López M, Tarrés M, Ibars E, Solé-Soler R, Gutiérrez-Escribano P, Apostolova S, Casas C, *et al* (2019) Sumoylation of Smc5 Promotes Error-free Bypass at Damaged Replication Forks. *Cell Reports* 29: 3160-3172.e4

Zhang Y, Hou K, Tong J, Zhang H, Xiong M, Liu J & Jia S (2023) The Altered Functions of Shelterin Components in ALT Cells. *International Journal of Molecular Sciences* 24: 16830

Zhao Q, Ma Y, Li Z, Zhang K, Zheng M & Zhang S (2020) The Function of SUMOylation and Its Role in the Development of Cancer Cells under Stress Conditions: A Systematic Review. *Stem Cells Int* 2020: 8835714

Zhao X & Blobel G (2005) A SUMO ligase is part of a nuclear multiprotein complex that affects DNA repair and chromosomal organization. *Proc Natl Acad Sci U S A* 102: 4777–4782

Zhou B-BS & Elledge SJ (2000) The DNA damage response: putting checkpoints in perspective. *Nature* 408: 433–439

Zhou J, Wu G, Tong Z, Sun J, Su J, Cao Z, Luo Y & Wang W (2020) Prognostic relevance of SMC family gene expression in human sarcoma. *Aging (Albany NY)* 13: 1473–1487

Zhuang J, Shirazi F, Singh RK, Kuiatse I, Wang H, Lee HC, Berkova Z, Berger A, Hyer M, Chattopadhyay N, Syed S, Shi JQ, Yu J, Shinde V, Tirrell S, Jones RJ, Wang Z, Davis RE, Orlowski RZ. (2019) Ubiquitin-activating enzyme inhibition induces an unfolded protein response and overcomes drug resistance in myeloma. Blood. Apr 4;133(14):1572-1584.

Appendix- Manuscript #4 (Co-Author)

The Smc5/6 core complex is a structure-specific DNA binding and compacting machine

Diego Serrano[1,*], Gustavo Cordero[2,#], Ryo Kawamura[3,#], Aleksandr Sverzhinsky[4,#], Muzaddid Sarker[1], Shamayita Roy[1], Catherine Malo[1], John M. Pascal[4], John F. Marko[3] and Damien D'Amours[1,5,†]

Published in : Molecular Cell 2020 Dec 17;80(6):1025-1038.e5.
doi:10.1016/j.molcel.2020.11.011. Epub 2020 Dec 9

[1] Ottawa Institute of Systems Biology, Department of Cellular and Molecular Medicine, University of Ottawa, Roger Guindon Hall, 451 Smyth Rd, Ottawa, ON, K1H 8M5, Canada
[2] Département de pathologie & biologie cellulaire, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, QC, H3C 3J7, Canada
[3] Department of Molecular Biosciences and Department of Physics & Astronomy, Northwestern University, Evanston, IL, 60208, USA
[4] Département de biochimie et médecine moléculaire, Université de Montréal, Montréal, Québec, H3C 3J7, Canada
* Current address: Department of Pathology, Anatomy and Physiology, Faculty of Medicine – Solid Tumors Program, Center for Medical Applied Research; University of Navarra, Avenida Pio XII 55, 31008, Pamplona, Spain
[#] These authors contributed equally to this work.

[5] Lead Contact: Damien D'Amours
[†] *Correspondence: damien.damours@uottawa.ca*

DS, RK, AS, JMP, JFM & DD conceived and designed the experiments; DS performed the *in vivo* fusion and XL-MS experiments; DS, MS & GC performed purifications, DNA-binding & enzymatic assays; GC carried out the analytical ultracentrifugation & GraFix procedures; AS & JMP performed EM imaging; RK carried out magnetic tweezers experiments; DS, **SR** & CM created yeast strains, assessed their viability & Smc6 abundance; DS, GC, RK, AS, JMP, JFM & DD analyzed the data; DS, RK, **SR** & AS prepared figures; DS & DD wrote the paper.

Summary

The structural organization of chromosomes is a crucial feature that defines the functional state of genes and genomes. The extent of structural changes experienced by genomes of eukaryotic cells can be dramatic and spans several orders of magnitude. At the core of these changes lies a unique group of ATPases –the SMC proteins– that act as major effectors of chromosome behavior in cells. The Smc5/6 proteins play essential roles in the maintenance of genome stability, yet their mode of action is not fully understood. Here we show that the human Smc5/6 complex recognizes unusual DNA configurations and uses the energy of ATP hydrolysis to promote their compaction. Structural analyses reveal subunit interfaces responsible for the functionality of the Smc5/6 complex and how mutations in these regions may lead to chromosome breakage syndromes in humans. Collectively, our results suggest that the Smc5/6 complex promotes genome stability as a DNA micro-compaction machine.

Introduction

Maintenance of genome integrity is an essential function for all living organisms. We know that several pathways act simultaneously *in vivo* to ensure the maintenance of genome integrity, but how these multiple pathways successfully integrate and/or coexist in time and space remains an outstanding question. Importantly, the ability to integrate disparate processes into an effective cellular response likely depends on the action of higher-order effectors that can modulate the physical conditions of the genome to allow multiple molecular transactions to operate simultaneously at a given genomic location.

The structural maintenance of chromosomes (SMC) family of proteins represents one class of higher-order effectors of genome organization that can coordinately regulate multiple processes *in vivo* (reviewed in Aragon, 2018; Baxter et al., 2019; Yatskevich et al., 2019). These proteins are mechanochemical regulators of chromatin behavior and their activity is essential for the protection and dissemination of genomes in all species. SMC proteins assemble into large ring-like complexes that include cohesin (Smc1/3), condensin (Smc2/4) and the more recently discovered Smc5/6 complex. In eukaryotes, condensin participates in the compaction of chromosomes during cell division, whereas cohesin enforces proximity and alignment of sister chromatids (Yatskevich et al., 2019). The Smc5/6 complex, in contrast, plays diverse roles in the cellular response to DNA damage, but its exact contribution to genome stability is still unclear (reviewed in Aragon, 2018). A remarkable feature of the Smc5/6 complex is that, despite its crucial contribution to DNA repair, segregation and replication (Irmisch et al., 2009; Jeppsson et al., 2014; Potts et al., 2006), it is not a core factor essential for these individual processes under standard conditions. Instead, it appears to provide an integrative function that prevents toxic interactions involving distinct molecular pathways (Chen et al., 2009; Pebernard et al., 2006; Torres-Rosell et al., 2005).

How the Smc5/6 complex promotes chromosome integrity is a key unanswered question in the genome stability field. This gap in our knowledge can be explained in part by the difficulty in isolating pure Smc5/6 complex to perform in-depth mechanistic and structural studies. We report herein the development of an innovative co-translational folding strategy to purify the human Smc5/6 complex in its active form. Using this approach, we have discovered that the human Smc5/6 complex is a structure-specific DNA binding and compacting machine. An accompanying study (Gutierrez-Escribano et al., 2020) has found identical activities in the budding yeast complex, underscoring the conservation in the mode of action of eukaryotic Smc5/6 enzymes.

Results

A co-translational folding approach to assemble the human Smc5/6 complex

Previous studies from our laboratory and others have shown that the yeast Smc5-Smc6 dimer (Roy et al., 2015) as well as a number of human non-SMC elements of the complex (Nse1-Nse3; Doyle et al., 2010; Zabrady et al., 2016) can be purified as recombinant proteins. However, our initial efforts (as well as that of others; Zabrady et al., 2016) to purify soluble Nse4 monomer or recombinant Smc5/6 holoenzyme have been unsuccessful (Fig. 5.S1A). This is likely due to the high number of hydrophobic clusters present in Nse4 (Fig. 5.S1B), a central interaction hub within the Smc5/6 complex (Fig. 5.1A).

We considered the possibility that enforcing co-translational folding of hNse4a with other members of the Smc5/6 holoenzyme might facilitate complex assembly *in vivo* (Fig. 5.1A). To achieve this goal, we used a flexible linker to physically connect Nse4a with either Smc5 or Smc6, similar to a strategy previously employed to covalently close the cohesin ring (Gruber et al., 2006). Multiple TEV cleavage sites were introduced in the linker sequence between Nse4a and Smc5 or Smc6 to facilitate

linker removal and recovery of individual subunits with native (or near-native) sequence after purification. We created two different yeast strains overexpressing Nse1 and Nse3 together with Nse4a-Linker-Smc5 or Smc6-Linker-Nse4a and their respective SMC partners (Fig. 5.1B). The hNse2 protein was not included in these complexes because it dissociates from the holoenzyme during mitosis (Behlke-Steinert et al., 2009) and studies suggest that a non-negligible fraction of Smc5/6 complexes do not contain Nse2 (see gel filtration fractions 10 and 11 in Andrews et al., 2005). Consistent with this view, Nse2 is not essential for cellular viability in metazoans (*i.e.*, human, mouse and chicken cells; Kliszczak et al., 2012; Verver et al., 2016; Zheng et al., 2017), whereas all other components of the complex are (Aragon, 2018). The human complex containing Smc5/6 and Nse4/3/1 subunits corresponds to the canonical organization of other SMC complexes (condensin and cohesin; reviewed in Yatskevich et al., 2019), and we will hereafter refer to this form of the enzyme as the Smc5/6 core complex. As a control, we constructed an additional overexpression strain specific for the SMC scaffold of the enzyme, the Smc5/6 dimer.

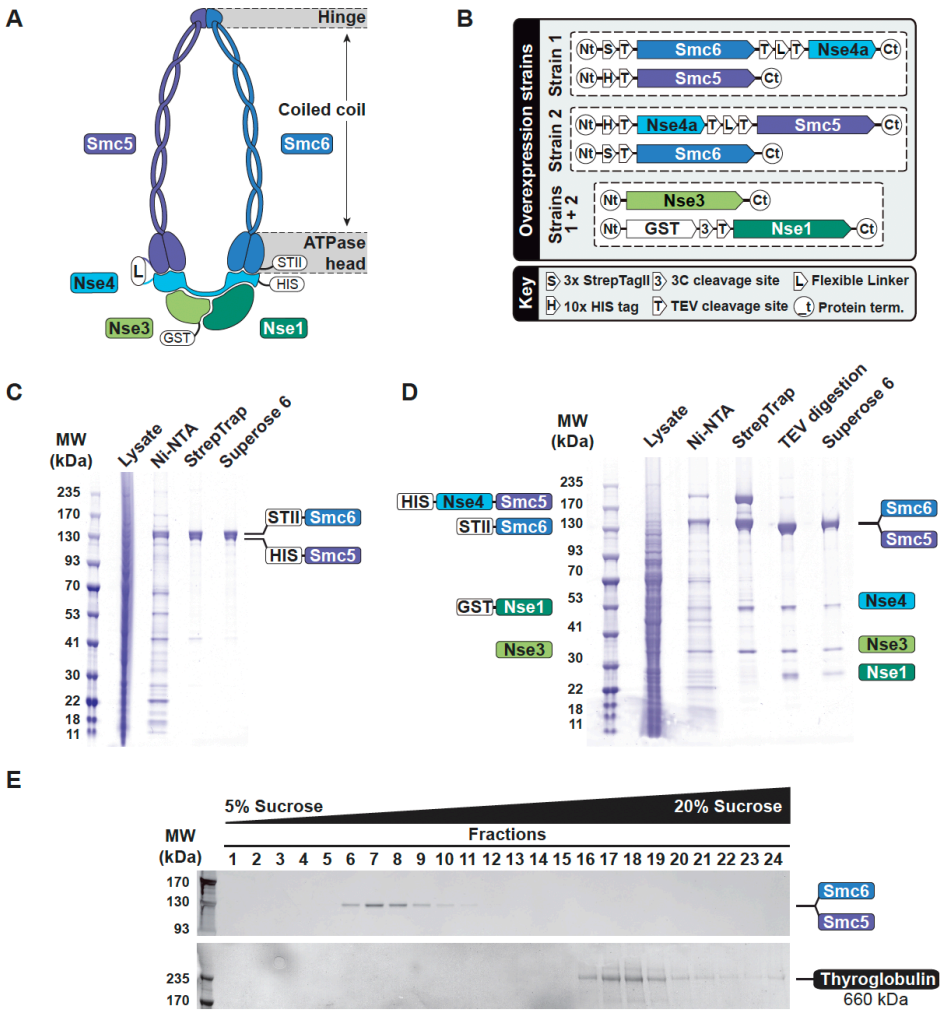


Figure 5.1. Purification of the Smc5/6 complex. (A) Schematic representation of the Smc5/6 complex. **(B)** Overview of the 2 yeast strains used for Smc5/6 complex overexpression and purification. Each strain contains 2 plasmids; one for the overexpression of the Nse4/SMC subunits and the other –common to both strains– that overexpresses Nse1 and Nse3. **(C)** Representative gel showing the different chromatography steps used to purify the Smc5/6 dimer. The positions of StreptagII-Smc6 and His-Smc5 are shown on the right. **(D)** Purification steps used to isolate the Smc5/6 core complex. The positions of individual proteins prior to TEV cleavage are shown on the left of the gel, while the positions of subunits after TEV-induced removal of purification tags/linker are shown on the right of the gel. After completion of the TEV cleavage reaction, all the subunits of the Smc5/6 complex migrate in SDS-PAGE at the positions of the native full-length proteins. Note that tag-less Nse4a migrates at the same apparent mass as GST-Nse1 (*i.e.*, prior to TEV-induced removal of the GST tag). Likewise, TEV and tag-less Nse1 are co-migrating on the gel, but TEV is removed from the Smc5/6 complex at the final gel exclusion step. **(E)** Analysis of the oligomerization status of the Smc5/6 core complex by density gradient centrifugation. See also Figures 5.S1-3.

Purification of functional Smc5/6 dimer and core complex

We took advantage of a yeast co-overexpression system developed in our laboratory (Roy et al., 2011; St-Pierre et al., 2009) to isolate the Smc5/6 heterodimer to near homogeneity (Fig. 5.1C). Importantly, the stoichiometry of the SMC subunits was 1:1 after our purification, reflecting the stability of subunits and proper dimerization through the hinge domain. Small purification tags were not removed from Smc5 and Smc6 since previous work from our laboratory showed that the Smc5/6 dimer is not adversely affected by fusion with multi-histidine and Strep-tagII sequences (Roy and D'Amours, 2011; Roy et al., 2011).

We next proceeded to purify the Smc5/6 core complex using the fusion approach described above. We focused our initial efforts on complexes containing the Nse4-Linker-Smc5 fusion protein (Fig. 5.1B). We were able to purify soluble recombinant Smc5/6 core complex using a combination of nickel-chelate (Ni-NTA), StrepTactin and glutathione affinity chromatography steps (Figs. 5.1D, 5.S2A). After cleaving the purification tags (on Nse1, Nse4a and Smc6) and linker sequence connecting Nse4 and Smc5, we performed size exclusion chromatography on Superose 6 as a final step to remove TEV protease and tag remnants (Figs. 5.1D, 5.S2B). This co-translational expression/folding strategy enabled the purification of soluble Smc5/6 core complex to high levels and near homogeneity (≥95%). Rate-zonal centrifugation of the purified Smc5/6 core complex revealed its subunits sedimented in fractions corresponding to a mass of 300-400 kDa, a range that is consistent with a monomeric complex (*i.e.*, predicted mass = 365 kDa; Figs. 5.1E, 5.S3A,B).

The Nse4-Smc5 fusion is functional in vivo

We next wanted to determine whether fusion of Nse4 with SMC subunits is compatible with assembly and full functionality of the Smc5/6 complex *in vivo*. To test this possibility, we expressed Nse4-Linker-

Smc5 and Smc6-Linker-Nse4 fusion proteins in the budding yeast *Saccharomyces cerevisiae*. Sporulation of heterozygous diploid strains carrying *NSE4/nse4Δ SMC5/smc5Δ* showed that expression of the yeast *NSE4-Linker-SMC5* fusion protein can complement cell lethality in spores carrying both *smc5Δ* and *nse4Δ* (Fig. 5.2A). Moreover, *smc5Δ nse4Δ* mutants expressing the yeast Nse4-Linker-Smc5 protein showed wild-type kinetics of cell proliferation in the presence or absence of DNA damaging agents (Fig. 5.2B-C). Since the Smc5/6 complex is required for cell viability and DNA repair, we infer from these results that fusing Nse4 to Smc5 does not adversely affect the functionality of the complex.

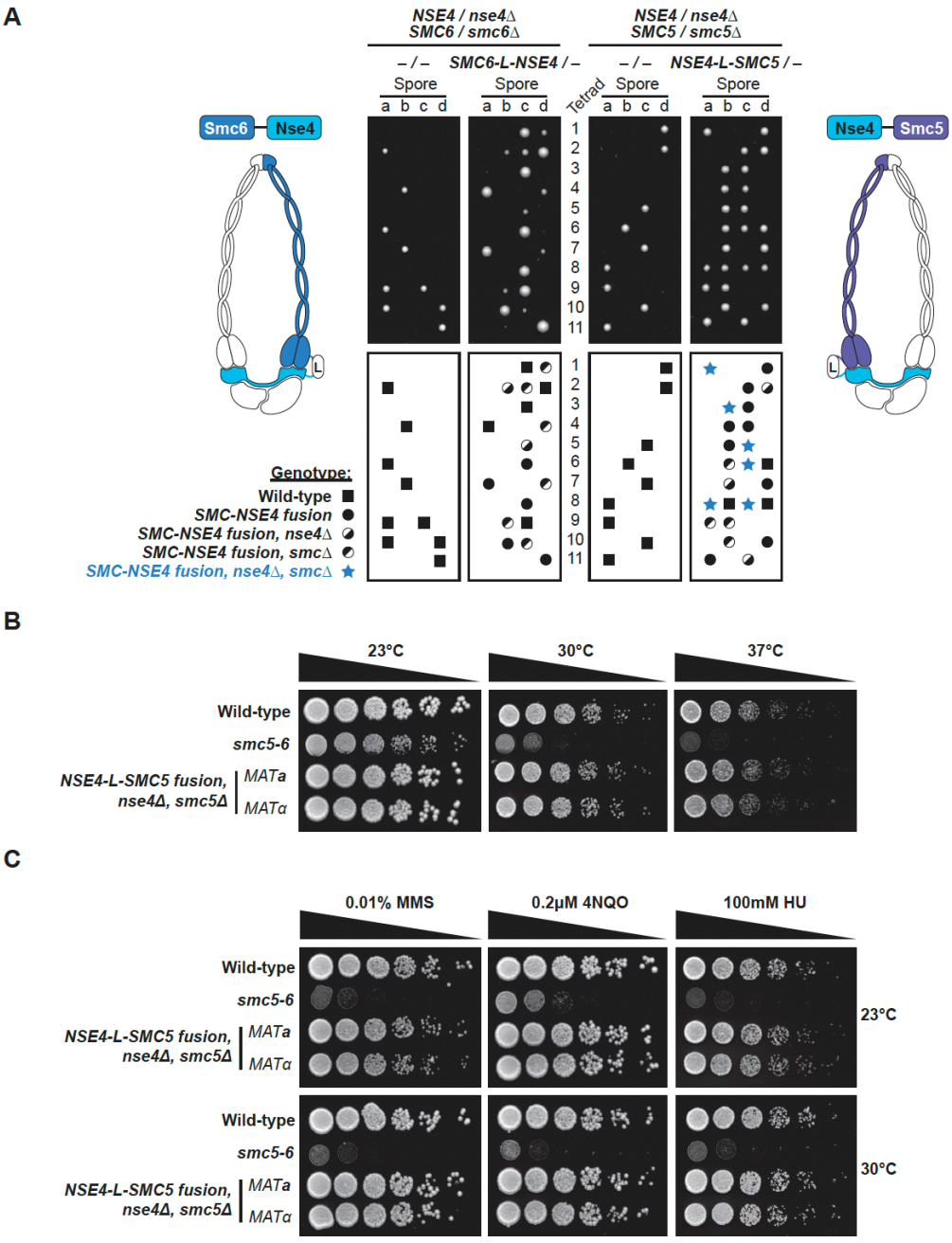


Figure 5.2. Functionality of Smc5/6 complexes containing Nse4-SMC fusions. (A) Phenotype of *SMC5/6-NSE4* fusion alleles after sporulation. One copy of *NSE4* and either *SMC5* or *SMC6* were deleted in diploid yeast strains expressing Nse4-linker-Smc5 or Smc6-linker-Nse4 fusion proteins. The linker sequence and fusion strategy employed in these strains are identical to those used to purify the human subunits, except that yeast Smc5, Smc6 and Nse4 protein sequences were used to allow complementation of *nse4Δ*, *smc5Δ* and *smc6Δ* deletions. Diploid strains of the indicated genotype were sporulated and haploid spores micromanipulated on solid growth medium. The viability of spores was scored after 3 days of germination, and their genotypes are represented schematically under the growth plates. –/– and *NSE4-L-SMC*/– indicate the absence or presence of fusion alleles at

the *URA3* locus of parental/diploid yeast. **(B-C)** Proliferation capacity of two independent *smc5Δ nse4Δ* clones (*MAT**α*** and *MAT**α***) expressing the Nse4-linker-Smc5 fusion protein was assessed by serial dilution assay on solid medium. After plating, cells were grown for 2-3 days at the indicated temperatures (23 °C, 30 °C and 37 °C; panel B) in the absence or presence of DNA damaging drugs (HU, MMS or 4NQO; panel C).

Arm configuration of the Smc5/6 complex

We took advantage of our purified complex to visualize the general morphology and arrangement of Smc5/6 subunits by negative stain transmission electron microscopy (EM). To ensure the structural integrity of Smc5/6 complexes used in our EM experiments, we first fractionated the enzyme by sucrose density centrifugation and subsequently imaged a single fraction corresponding to monodisperse core complexes (*i.e.*, fraction 7; Figs. 5.1E, 5.3A). Image analysis revealed that particles corresponding to the entire Smc5/6 complex adopted a rod conformation rather than a ring shape (Fig. 5.3B,E). Since the majority of observed particles were smaller than expected, we decided to use mild chemical crosslinking (GraFix) to rule out the possibility that the conformation observed for the Smc5/6 complex was due to an artifact of grid preparation and/or complex dissociation. The GraFix procedure stabilizes the structure of protein complexes by virtue of progressive fixation in a sucrose gradient containing a low concentration of glutaraldehyde (Fig. 5.S3C)(Kastner et al., 2008; Stark, 2010). Upon completion of GraFix ultracentrifugation, fractions corresponding to the pentamer (*i.e.*, fraction 7; compare Figs. 5.1E/5.S3A and 5.3A/5.S3C) migrated as a single high-molecular mass band on SDS-PAGE gels, indicating efficient intra-complex crosslinking without inter-complex aggregation (Fig. 5.3A).

We then proceeded to EM image acquisition after negative staining of the stabilized complex. As expected, the GraFix samples showed an increased numbers of SMC-like particles per micrograph. Importantly, image analysis confirmed that the Smc5/6 core complex adopted a closed arm configuration with two globular domains (corresponding to the hinge domain, on top; and the

Nse1/3/4 subcomplex, bottom) linked by a rod structure (corresponding to the coiled-coil [CC] arms of Smc5/6; Fig. 5.3C,F). This morphology is visually analogous to that of two cherries connected to a single stem. Dimensions of the complex are slightly smaller than those of condensin and cohesin (namely ~450 x 140 Å), consistent with the shorter CC arms of Smc5 and Smc6 proteins (as predicted by Burmann et al., 2017; Fig. 5.3B-C,E-F). The resultant three-dimensional EM map of the Smc5/6 complex revealed that the size and shape of the globular regions on each end of the rod-shaped complex dovetail nicely with the dimensions of the Nse1/3 dimer and yeast Smc5/6 hinge domain observed in crystal structures (Fig. 5.S4A-F; Alt et al., 2017; Doyle et al., 2010). These observations suggest that the morphology reported here reflects a native configuration of the Smc5/6 complex.

The Smc5/6 dimer also appeared as a rod shaped-particle by negative stain EM, but we noticed that the SMC arms demonstrated more flexibility in the dimer than in the core complex (Fig. 5.3D,G). In some cases, it was possible to distinguish individual CC arms or kinks in the arms of the dimer in single particle images. This apparent increase in the flexibility of SMC arms is consistent with the absence of kleisin subunit in the Smc5/6 dimer, a condition that is expected to relieve structural constraints normally present on SMC arm movement. Taken together, these experiments reveal that subunits of the Smc5/6 core complex assemble into a rod-like structure that favors a colinear/closed arm configuration for the SMC subunits.

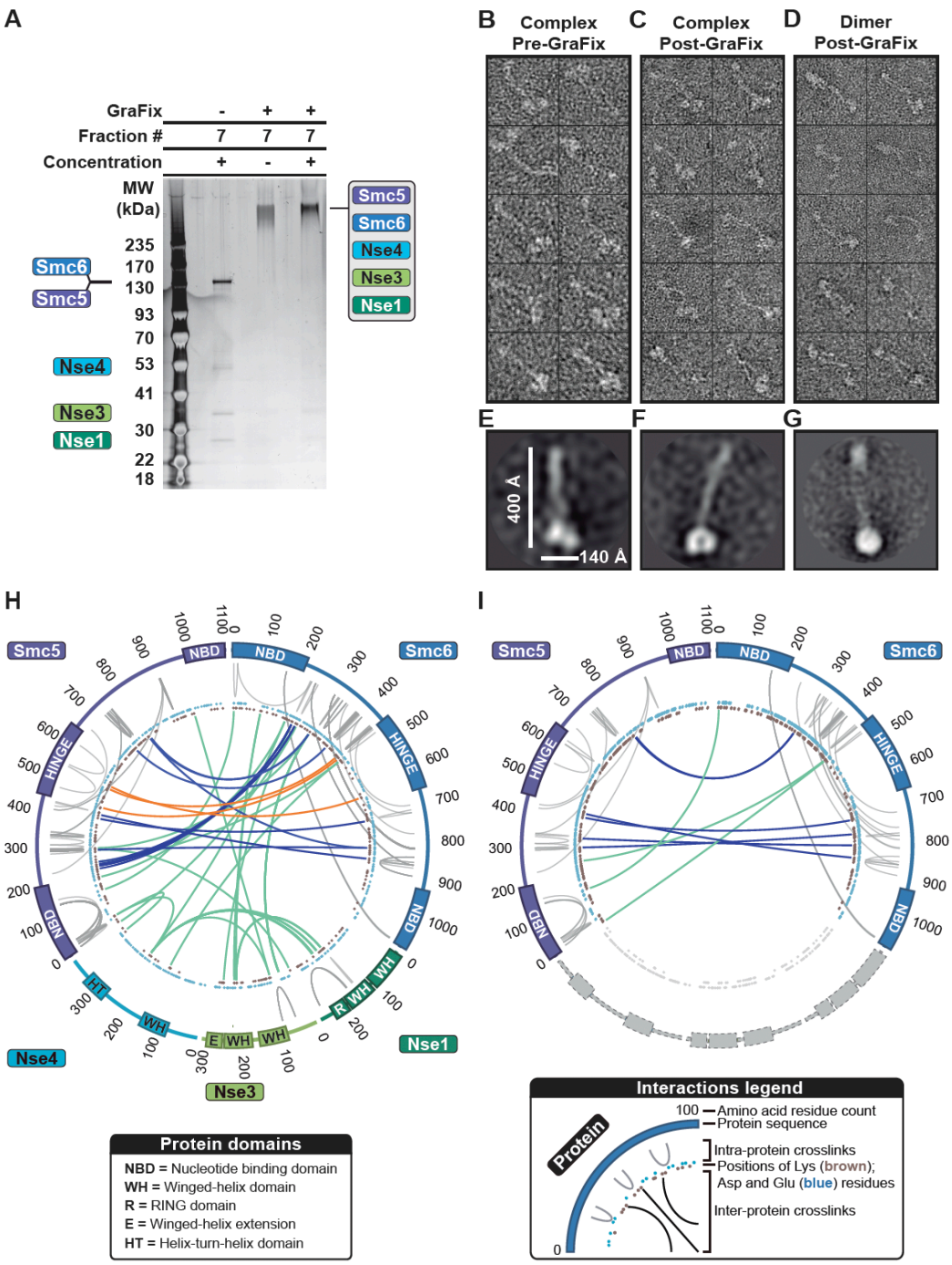


Figure 5.3. Shape and configuration of the Smc5/6 dimer and core complex. (A) Stabilization of the Smc5/6 core complex using the GraFix procedure (Kastner et al., 2008). In the absence of crosslinking, subunits of the complex eluted as individual bands in sucrose density gradients, whereas they eluted as a single band of high molecular mass after GraFix crosslinking. **(B-C)** Representative single particles of Smc5/6 core complexes prior (B) and after (C) GraFix treatment. Box dimensions are 518 Å x 518 Å. **(D)** Single particle images showing the Smc5-6 heterodimer after GraFix stabilization. **(E-F)** Two-

dimensional class averages of Smc5/6 core particles pre- (E) and post-GraFix treatment (F). Dimensions are represented in Å and images in E and F were reconstituted from 44 and 68 particles, respectively. **(G)** Two-dimensional class average of GraFix-stabilized Smc5/6 dimer (n = 167 particles). **(H)** Proximity maps showing intra- and inter-subunit connections within the Smc5/6 core complex. XLs identified in the MS analysis were plotted on circular diagrams corresponding to the amino-acid sequences/functional domains of Smc5/6 complex subunits. Intra-molecular connections are shown as grey lines, whereas inter-molecular contact points are shown in the inner part of the diagram (blue lines are for XLs specific to CC domains; orange lines are for XLs connecting hinge domains; green lines are for all other XLs). **(I)** Network of inter- and intra-subunit connection points within the Smc5/6 dimer. See also Figures 5.S3-5.S4 and Table 5.S1.

Mapping contact points among the subunits of the Smc5/6 complex

We used small chemical probes to explore the landscape of contact points among the subunits of the Smc5/6 complex. The purified dimer and core complex were exposed to agents that crosslink (XL) lysines to other residues and reactions were subsequently processed by mass spectrometry (MS), as performed previously (Courcelles et al., 2017). In this analysis, XLs are formed between different amino-acid residues based on their proximity within the Smc5/6 complex, thereby indicating connection points or transient interfaces among subunits.

Graphical maps displaying the network of subunit connections identified by XL-MS are shown in Figure 5.3H-I, whereas the complete list of 158 intra- and inter-subunit connections identified in our experiments is included in Table S1. Analysis of the landscape of connections in these maps confirmed that the dimer is highly flexible in solution, which is demonstrated in proximity maps as a low frequency of interaction for residues of different SMC subunits (Table 5.S1, 9 XLs; Fig. 5.3I, colored lines). In comparison, when the movement of SMC arms is constrained by the kleisin subunit in the core complex, connections involving different SMC proteins were more than twice as abundant (Table 5.S1, 20 XLs; Fig. 5.3H). These observations are consistent with the level of SMC arm flexibility that we noticed in EM experiments (Fig. 5.3B-G).

A large fraction (~60%) of the long-range connections that we identified are consistent

with the known structure of SMC complexes, thereby validating the quality of the XL-MS results. For example, the XL-MS analysis revealed several connections involving the hinge domains of the two SMC subunits (*e.g*., Smc5-K542 with Smc6-K628; Fig. 5.3H, orange lines). We also detected several predicted intra-molecular connections among the α-helical regions that form the CC arms of individual SMCs (Fig. 5.S4G,H). Our EM data showed that the SMC arms of Smc5 and Smc6 were in close proximity to each other in the holoenzyme, an observation independently confirmed by our XL-MS data (see inter-subunit XL connecting Smc5/6 CC arms; Smc5-K232/K247/K300/K783 with Smc6-K234/248/K310/K821, respectively; Figs. 5.3H, 5.S4I).

Beyond the XLs connecting SMC proteins together, our analysis revealed a rich network of connections linking functional domains of NSE subunits with other parts of the complex, such as the winged-helix (WH) domains of Nse3 and the N-terminal ATPase domain/CC neck of Smc6 (Fig. 5.3H). The α helical region of Smc6 acted as a connection hub for all but one subunit of the core complex (*i.e*., Nse1), suggesting an important role in the regulation of complex activity or topology. We also detected several contact points involving Nse3 to Nse1, as predicted from the structure of the dimer (Doyle et al., 2010). Our XL-MS data suggest that Nse4 in combination with Nse1 and Nse3 is capable of bridging the ATPase/CC neck regions of Smc5 and Smc6.

Functional relevance of contact points in Smc6 and Nse3-4 subunits

Next, we investigated the physiological significance of contact points that we identified in the WH domain of Nse3 and other parts of the complex. This region is of particular interest because it contains mutations responsible for the lung disease, immunodeficiency, and chromosome breakage (LIC) syndrome (Fig. 5.4A-C; van der Crabben et al., 2016). Based on the evolutionary conservation of the crosslinked positions and nearby residues, we inserted several mutations in the *NSE3* locus in *S.*

cerevisiae (Fig. 5.4B,C) and assessed the ability of the resulting mutant strains to proliferate in the presence of DNA damaging agents (Fig. 5.4D).

Yeast strains carrying mutations in the WH-B extension domain of Nse3 demonstrated a strong conditional defect in their ability to grow on media containing genotoxic agents like MMS, HU or 4NQO (Fig. 5.4D). Whereas the proliferation of *nse3-E265R* and *nse3-L268K* mutant appeared normal on media containing DNA damaging agents at 23 °C, the growth capacity of these strains became severely affected at 37 °C (Fig. 5.4D). In fact, yeast strains carrying the *nse3-E265R* allele were sensitive to all DNA damaging agents tested, whereas the growth of *nse3-L268K* mutants was strongly impaired in the presence of MMS at 37 °C. Importantly, the *nse3-E265R* and *nse3-L268K* mutations map to a highly conserved α helix in the WH-B extension domain and are located very close (*i.e.*, less than 12 Å and 6 Å, respectively) to a residue mutated in LIC syndrome patients (Fig. 5.4C, inset). From a functional perspective, the corresponding hNse3 residues represent contact points with the N-terminal ATPase domain of Smc6 in our XL-MS data (Fig. 5.3H). We also noticed that the same region of Smc6 (together with the adjacent CC neck) shared reciprocal connections with other residues in the WH domain of hNse3. This suggests the existence of an interaction hub for Nse3 in the ATPase/CC neck of Smc6.

To test the idea proposed above, we introduced mutations in residues of yeast Smc6 that correspond to those interacting with the WH domain of Nse3 (*i.e.*, positions marked in Fig. 5.S4G). As predicted, yeast expressing the *smc6-R135E* and *smc6-R144E* alleles showed severe sensitivity to DNA damaging agents at 37 °C (Fig. 5.4D). We also constructed a yeast strain carrying a D261A mutation in Nse4, another putative Smc6-proximal residue. The *nse4-D261A* mutants showed growth defects highly reminiscent of those seen in *smc6* and *nse3* mutants (Fig. 5.4D), underscoring their common hypomorphic defect. Remarkably, mutation in arginine 144 of Smc6 led to a substantial decrease in

protein abundance at non-permissive temperature (Fig. 5.4E-F), a phenotype similar to the protein stability defect observed in subunits of the Smc5/6 complex of LIC syndrome patients (van der Crabben et al., 2016). Together, these results indicate that residues at the interface of Nse3-4 and the ATPase/CC neck region of Smc6 are important for the activity of the Smc5/6 complex.

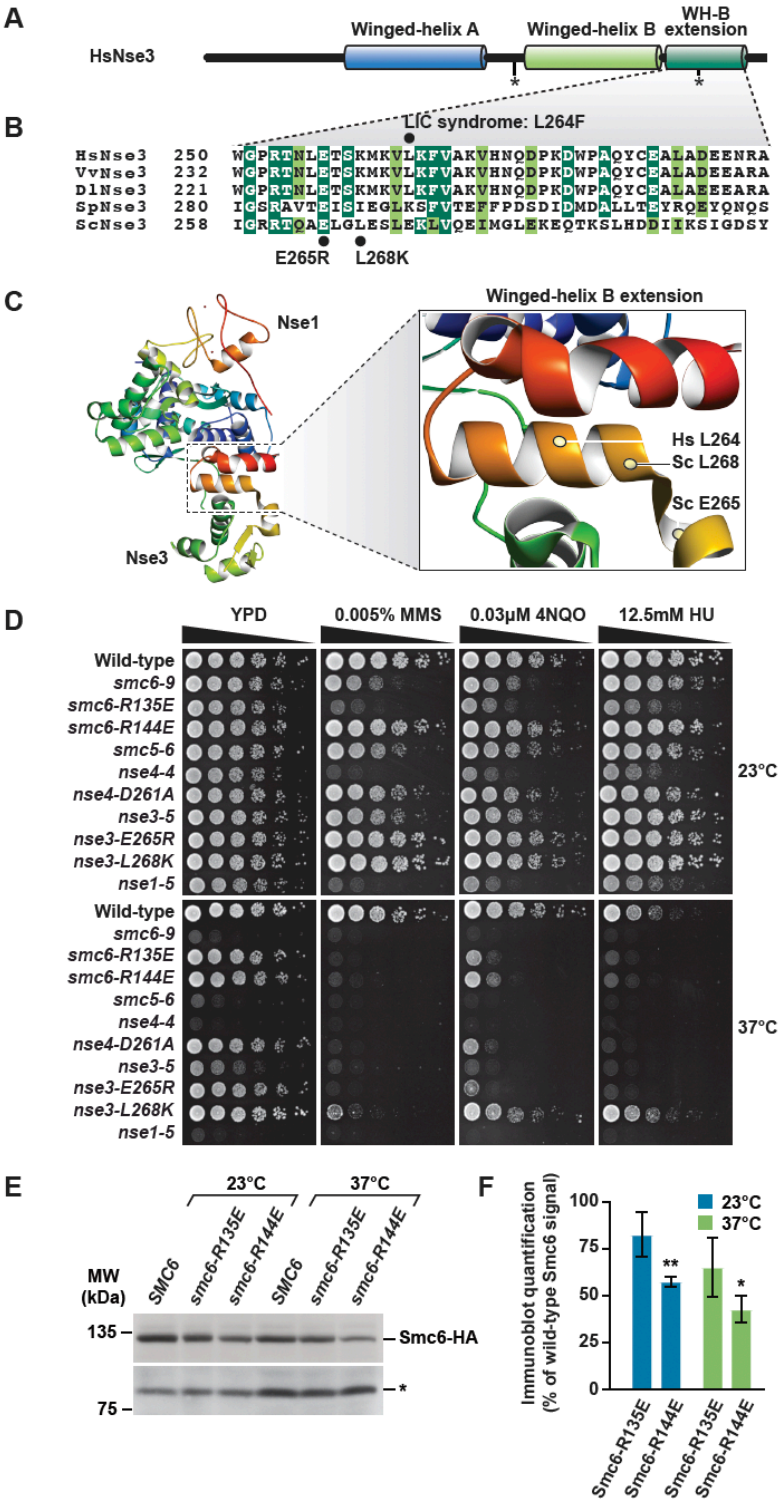


Figure 5.4. Impact of mutations in the WH domain of Nse3 and the ATPase/CC neck of Smc6. **(A)** Schematic representation of Nse3 domain structure. The positions of LIC syndrome mutations are marked with asterisks (van der Crabben et al., 2016). **(B)** Alignment of the WH-B extension domain in eukaryotic homologs of Nse3. **(C)** Crystal structure of the Nse3-Nse1 dimer. The position of mutations

created in this study is marked with yellow ovals in the magnified view (inset). The structure is PDB 3NW0 from Newman et al. (2016). **(D)** Proliferation phenotype of yeast strains carrying mutations in Smc5/6 complex components. Strains were diluted on solid medium containing DNA damaging agents and grown at the indicated temperatures for 2-4 days before recording their phenotype. **(E)** Effect of temperature on the stability of Smc6 mutants. Cell lysates of yeast grown at the indicated temperatures were resolved by SDS-PAGE and processed for immunoblot analysis. The positions of Smc6 and loading control (*) bands are shown on the right. **(F)** Quantitative analysis of Smc6 abundance determined by immunoblot. The bar graph reports the mean protein abundance of Smc6-R135E/R144E relative to wild-type Smc6 ± SEM for 3 independent experiments. * signifies *p*-value ≤ 0.05 and ** a *p*-value ≤ 0.01 (ANOVA with Dunnett's *post hoc* test).

Structure-specific DNA recognition and binding by the Smc5/6 complex

Previous findings from our laboratory revealed that yeast Smc5 and Smc6, as well as the Smc5/6 dimer, have intrinsic DNA-binding activity (Roy and D'Amours, 2011; Roy et al., 2015; Roy et al., 2011). However, the substrate preference and affinity of the human Smc5/6 dimer and core complex remain to be established.

We focused our initial analysis on the Smc5/6 dimer to define the baseline DNA binding activity of these SMC proteins. We specifically monitored the affinity of Smc5/6 for double-stranded (ds) and single-stranded (ss) DNA substrates using electrophoretic mobility shift assays (EMSAs), as previously described (Roy et al., 2015). The nucleotide content of ss and dsDNA substrates used in this assay is roughly similar, thus allowing a direct comparison of Smc5/6 affinities for these substrates. Figure 5.5A shows that the Smc5/6 dimer exhibited different DNA binding behavior when ss and ds substrates were used. The affinity of the dimer was markedly higher for ssDNA substrates than for duplex DNA. For example, Smc5/6 could fully bind the ssDNA substrate at 48-fold molar excess, whereas equivalent saturation was only attained at 400-fold excess of dimer relative to the nucleotide content of the dsDNA (Fig. 5.5A).

Next, we asked if non-SMC elements (Nse4a/3/1) contribute to the DNA-binding properties of the Smc5/6 core complex. It has been shown previously that the human Nse1/Nse3 dimer can bind

various types of DNA (Zabrady et al., 2016), but whether this role can be performed by these subunits in the context of the Smc5/6 complex is still unknown. Our results indicate that the presence of the regulatory subunits does not alter the basal DNA-binding activity of the Smc5/6 dimer. As shown in Figure 5.5A-B, the affinity of the core complex for ssDNA is similar to the K_d of the dimer (32.4 nM ± 1.9 nM dimer, 27.3 nM ± 0.76 nM core complex; Fig. 5.S5A). From a qualitative standpoint, however, the nature of the binding by the core complex appeared different than that of the SMC dimer. The presence of the NSE elements in the core complex allowed the formation of intermediate DNA-protein species in the gel (*i.e.*, partially-retarded DNA; line with asterisk) which were not formed when the SMC dimer associated with ssDNA (Fig. 5.5A *vs* 5.5B; left) or were formed at a much lower abundance when the SMC dimer associated with dsDNA substrates (Fig. 5.5A *vs* 5.5B; center). Moreover, compared to the SMC dimer alone, NSE proteins lowered the dissociation constant of the core complex for dsDNA (121 nM ± 7 nM dimer *vs* 81.8 nM ± 3.04 nM core complex; Fig. 5.S5A).

We next sought to test whether the Smc5/6 complex is capable of binding small 60 bp DNA duplexes in EMSA experiments. We included RNA/DNA hybrids in this analysis since Smc5/6 was proposed to interact with this type of nucleic acid in the early stages of DNA replication (Lafuente-Barquero et al., 2017). We found that the Smc5/6 complex is able to bind both hybrid and dsDNA with similar affinity (Fig. 5.5C). RNA/DNA hybrids appeared to be slightly better binding partners for the Smc5/6 core complex than DNA duplexes, but this trend was not statistically significant. Together, our DNA-binding experiments indicate that the Smc5/6 complex can associate with non-B form DNA conformations with high affinity.

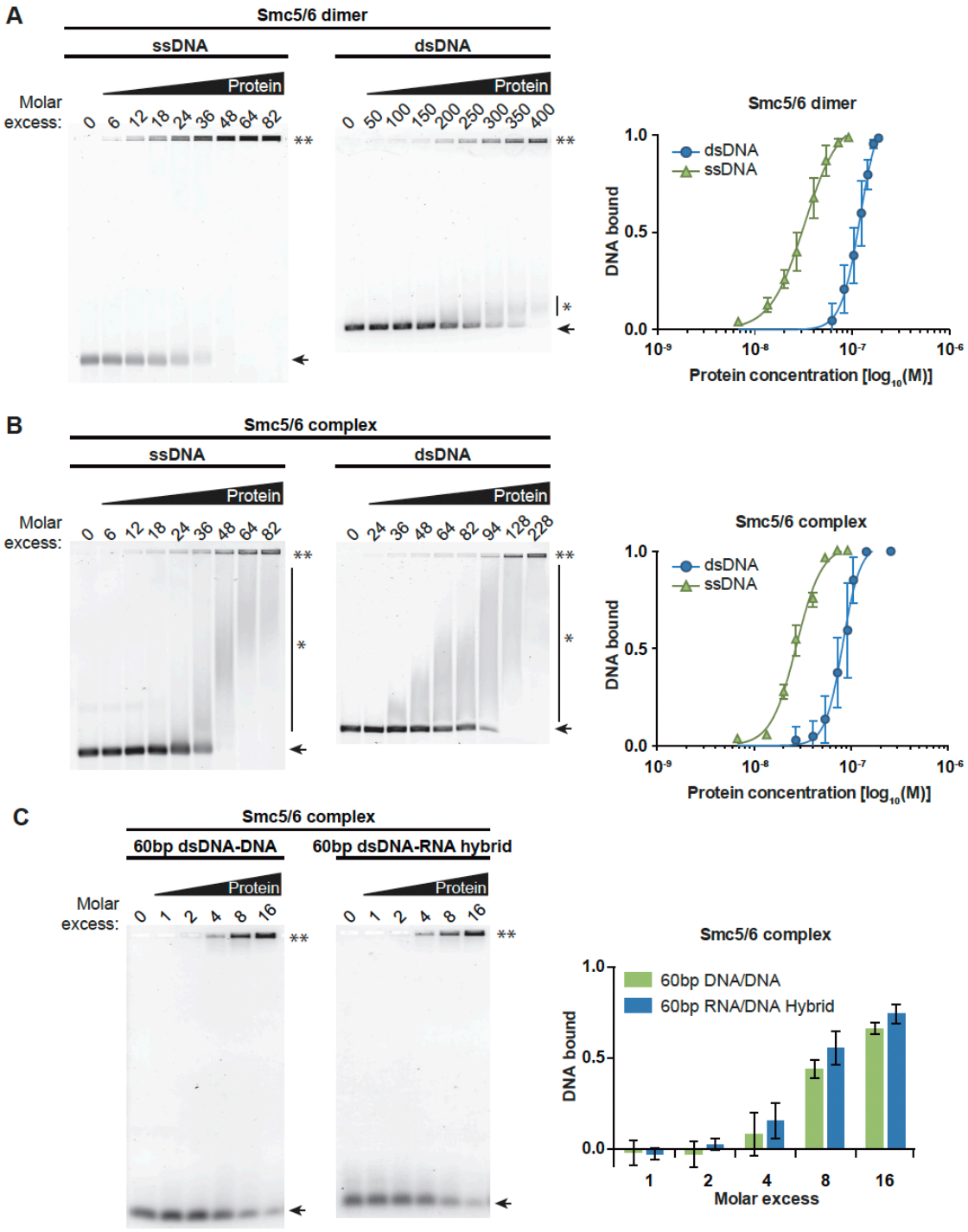


Figure 5.5. DNA-binding affinity and substrate preference of the Smc5/6 core complex. The DNA-binding behavior of the Smc5-6 dimer **(A)** and core complex **(B-C)** was determined by EMSA saturation experiments using various DNA substrates. Purified complexes were incubated with the indicated nucleic acids for 30 min at 30 °C and the resulting protein-DNA complexes were resolved by electrophoresis. The molar excess of Smc5/6 complex over DNA is shown above each lane, whereas the positions of unbound and Smc5/6-bound DNA substrates are marked by arrowheads and

asterisks, respectively. The graphs next to the agarose gels show the quantification of DNA-binding activity. The data reported in the graphs is the mean DNA binding ± SE of ≥3 independent experiments. See also Figure S5A.

The Smc5/6 core complex compacts DNA against force in an ATP-dependent manner

Next, we wanted to determine if the Smc5/6 complex was capable of remodeling the configuration of DNA in space. In order to monitor DNA compaction by the Smc5/6 core complex, we carried out single-DNA magnetic tweezers experiments (Keenholtz et al., 2017). This approach uses 10 kb end-labeled DNA molecules tethered to a cover glass at one end, and to a paramagnetic bead at the other end. Molecules were tethered inside flow cells, allowing solvent conditions in the vicinity of the DNA to be changed, and a nearby magnet was moved to vary the pulling force applied to the paramagnetic particle (Fig. 5.6A). A typical experiment started with DNA under 2 pN tension (a high enough force to pull nearly all bending fluctuations out of the DNA) in Smc5/6-free buffer. Then Smc5/6 core complex plus cofactors were introduced into the flow chamber, force was reduced to that under study, and the kinetics of the change in DNA extension were observed.

In the absence of Smc5/6 complex, DNA extension was stable for DNA tensions from 2 pN down to 0.3 pN. When 5 nM Smc5/6 core complex was added to the buffer in the absence of nucleotide, the DNA extension was stable and unchanged from its initial protein-free value as force was reduced (Fig. 5.6B; 5 nM Smc5/6 complex, 0 mM ATP). When the same experiment was performed with 1 mM ATP, DNA compaction was observed at a rate markedly dependent on force (Fig. 5.6C, 5 nM Smc5/6 complex, 1 mM ATP). Note that the compaction at 0.3 and 0.5 pN proceeds at a high rate and that the initial stages of the dynamics are in fact difficult to see in Figure 5.6C. At higher forces, individual steps are clearly visible (Fig. 5.6C, cyan 1 pN curve). Quantification of step size from a series of four 1 pN experiments led to an estimated step size of 90±10 nm, similar to that observed for condensin and cohesin complexes (Eeftens et al., 2017; Keenholtz et al., 2017; Strick et

al., 2004; Sun et al., 2013).

We next examined ATP-dependent compaction rates of the Smc5/6 core complex as a function of force (Fig. 5.6D). A strong suppression of compaction rate was observed when force increased above 0.5 pN, which is characteristic of a compaction reaction that depends on the capture of DNA loops (Keenholtz et al., 2017; Marko et al., 2019; Skoko et al., 2006; Sun et al., 2013). The total amount of compaction of the DNA (the ratio of the total change in extension during the compaction reaction to the initial extension) also showed a strong force dependence, again suppressed by forces ≥ 1 pN (Fig. 5.6E).

Given that the 0.5 pN reactions were in the midrange of compaction rate and total DNA compaction, we measured the amount of compaction at 0.5 pN for different nucleotide cofactors, relative to the case of 1 mM ATP (Fig. 5.6F, leftmost bar). When ATP was entirely omitted (2nd bar), DNA compaction was almost completely suppressed; the same effect was seen in the presence of 1 mM ADP (3rd bar) or when non-hydrolysable ATPγS was included in the reaction. These experiments are consistent with the fact that purified Smc5/6 enzyme is active as an ATPase in solution (below). We thus conclude that compaction of DNA against sub-piconewton forces by the Smc5/6 core complex requires nucleotide as well as nucleotide hydrolysis, and that the critical force that suppresses that compaction is approximately 1 pN.

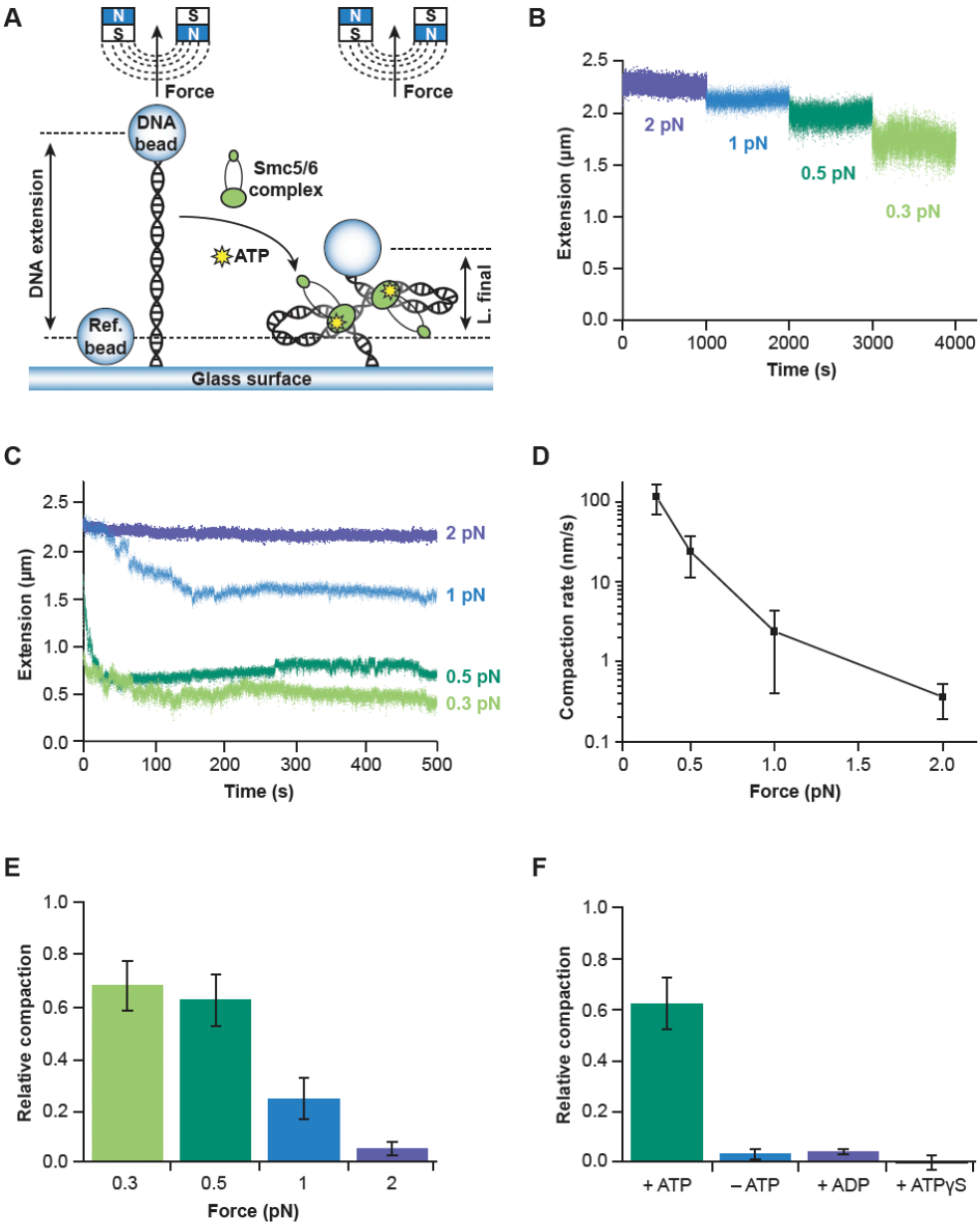


Figure 5.6. Compaction of DNA by the Smc5/6 complex. (A) Schematic representation of magnetic tweezers single-DNA compaction assay. **(B)** DNA extension measured following addition of 5 nM Smc5/6 complex (added slightly before t=0) with no nucleotide. As force was reduced from 2 pN to 0.3 pN, there was no compaction observed; the reduction of length is just that expected from thermal bending fluctuations of DNA, also responsible for the increased Brownian motion around the stable average DNA extension. **(C)** DNA extension following addition of 5 nM Smc5/6 complex (at t = 0) plus 1 mM ATP (separate curves are from separate experiments). At each force, compaction occurred, with more rapid and more complete compaction at lower forces. **(D)** Compaction rate in the presence of 1 mM ATP at 0.5 pN force measured from a series of 4 experiments of the type shown in panel C for each force (note logarithmic rate scale). A rapid increase in compaction rate with decreasing force

(essentially shut off above 1 pN) was observed, indicative of a DNA-loop-capture process of compaction. **(E)** Total compaction (ratio of change in extension to initial extension) in a series of 4 experiments at different forces, indicating that compaction is more complete at lower forces, and essentially shut off above 1 pN. **(F)** Total compaction at 0.5 pN in the presence of various nucleotides (0 and 1 mM ATP, 1 mM ADP, and ATPγS). All error bars indicate SEM.

Supercoiled DNA is a high affinity substrate for the Smc5/6 core complex

The observations that yeast and human Smc5/6 dimers can recognize structured DNA (Fig. 5.5; Roy and D'Amours, 2011; Roy et al., 2015; Roy et al., 2011) prompted us to investigate whether purified Smc5/6 core complex can preferentially bind to supercoiled DNA. We designed a "plectonemic supercoil capture" experiment (Fig. 5.7A), using the principle that when DNA is twisted while held under relatively low force (~0.3 pN), it will buckle when sufficiently twisted, and then form plectonemic supercoils, with each successive turn past the buckling point reducing the overall extension of the DNA molecule by about 50 nm (Strick et al., 1996). For the naked 10 kb DNAs used in this study, a linking number change (ΔLk) of +30 or -30 will induce a reduction in DNA extension of approximately 1200 nm; when ΔLk is returned to 0, this reduction in extension ceases because the DNA returns to its initial untwisted configuration.

To visualize the effect of the Smc5/6 complex on plectoneme dynamics, we used modified reaction conditions that do not allow the complex to compact DNA when ΔLk is set at 0 and low forces are applied on DNA (*i.e.*, no torsional stress; left purple points in Fig. 5.7B). Under these conditions, protein-DNA interactions are likely weakened to the point that the Smc5/6 complex was unable to capture loops of DNA (see Fig. 5.S5B for details). However, when we changed DNA linking number to ΔLk = -30 (Fig. 5.7B, blue points), waited for 30 seconds (Fig. 5.7B, green points), then returned to ΔLk = 0 (Fig. 5.7B, right purple points), DNA extension returned to a value about 1000 nm smaller than the starting point, consistent with supercoils being "captured" by the Smc5/6 core complex. We observed similar behavior for positive supercoiling (Fig. 5.7C). In the absence of ATP, we could

repeatedly cycle ΔLk between 0, +30, 0 and -30 without observing any "capture" events (Fig. 5.7D).

To characterize in detail the DNA supercoil-binding behavior of the Smc5/6 core complex, we carried out a series of single-molecule capture experiments where we specifically monitored the length of plectonemic DNA "stabilized" via supercoil binding/capture (Fig. 5.7A). The left group of bars in Figure 5.7E shows the reduction of length experienced by naked DNA after ΔLk of +30 and -30 (*i.e.*, step 2 in Fig. 5.7A). This reduction in length corresponds to the plectoneme size and is similar for the two directions of rotation; approximately 1200 nm. The central group of bars shows the reduction in length in the presence of 5 nM Smc5/6 core complex for ΔLk of +30 or -30 (*i.e.*, step 3 in Fig. 5.7A); again all the reductions in length are approximately 1200 nm indicating that the Smc5/6 core complex did little to perturb the extended length of plectonemic DNA formed under these conditions. However, the right group of bars shows that when ΔLk was returned to 0 (*i.e.*, step 4 in Fig. 5.7A), supercoils "captured" by Smc5/6 core complexes stabilized DNA in a plectonemic configuration (Fig. 5.7E). This "capture" effect was completely dependent on the presence of ATP in the reaction buffer, since DNA extension returned to its maximal/unfolded value –which reflects a plectoneme size of zero– in the absence of nucleotide. The amount of capture was on average equal for ΔLk of +30 or -30, with no apparent strong preference for one handedness or the other (Fig. 5.7E). The Smc5/6 core complex was capable of maintaining the captured DNA in a plectonemic configuration against forces up to 2 pN (Fig. 5.S5C). We conclude that under conditions where DNA is not compacted, the Smc5/6 core complex can efficiently bind and "capture" DNA supercoils in the presence of ATP, consistent with *in vivo* observations (Kegel et al., 2011). The effect of ATP in plectoneme stabilization experiments does not strictly imply a hydrolysis effect as it could also reflect ATP-mediated modulation of Smc5/6 complex binding to DNA.

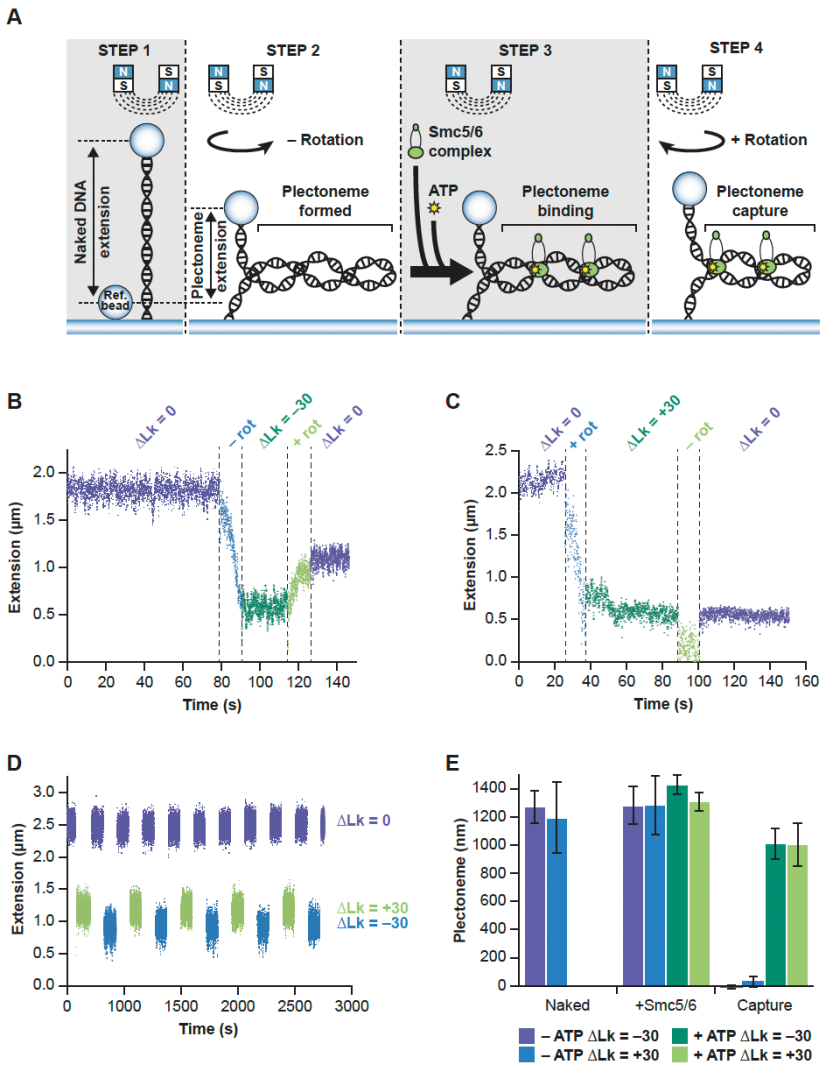


Figure 5.7. Capture of DNA by the Smc5/6 core complex requires supercoiling and ATP. (A) Schematic representation of the reaction steps in the plectoneme/supercoil capture experiment conducted under increased salt conditions. **(B)** At 100 mM potassium glutamate and 1 mM ATP, a 10 kb DNA under 0.3 pN force has stable extension (purple points, left); following increase of linking number to ΔLk = -30 (blue), the molecule forms plectonemic supercoils (dark green points); when linking number is returned to ΔLk = 0, only part of the DNA length contained within the plectoneme is returned (purple points, right); about 1000 nm of length has been "captured" by the Smc5/6 core complex. **(C)** Experiment similar to B showing capture of supercoiled DNA by the Smc5/6 core complex following supercoiling to ΔLk = +30 and return to ΔLk = 0. **(D)** In the absence of ATP, linking number can be reversibly and repeatedly cycled between ΔLk = 0, +30 and -30 with no capture of supercoiling by the Smc5/6 complex. **(E)** Averages of capture experiments over a series of 5 trials show that a large fraction of the DNA length change resulting from plectonemic supercoiling can be captured by the Smc5/6 core complex in the presence of ATP (*i.e.*, right bars; equivalent to step 4 in panel A). Changes observed in DNA extension/length are expressed as plectoneme size (nm) in the graph. The initial size of plectonemes formed with naked DNA (left bars; equivalent to step 2 in panel

A) or in the presence of 5 nM Smc5/6 core complex (central bars; step 3) are shown for comparison. Each bar shows mean and SEM. See also Figure 5. S5B,C.

Discussion

Compared to other SMC complexes, the Smc5/6 complex has remained an enigmatic effector of chromosome stability since its discovery. Whereas the core biological functions of cohesin and condensin are well defined, the involvement of the Smc5/6 complex in multiple, apparently distinct cellular processes has not allowed the emergence of a unified model to explain its unique contribution to genome stability (reviewed in Aragon, 2018).

Here we report for the first time how the human Smc5/6 complex recognizes and then modulates DNA structure in space. Our biophysical analyses suggest that the multi-functional nature of the Smc5/6 complex is achieved *in vivo* by compacting chromosomal regions containing unusual DNA structures. This compaction is focussed on specific DNA structures (*i.e*., local in nature), and we will thus refer to this process as "micro-compaction" to differentiate it from the global compaction seen in mitosis. We envision that the micro-compaction activity of the Smc5/6 complex will have at least two effects *in vivo*; first, it will create a physical barrier –steric hindrance– that will shield DNA intermediates from the action of undesirable modifying enzymes, and second, it will promote physical proximity of DNA molecules as a means to enhance biochemical reactions. The combined result of these two processes will be the stabilization and rapid repair of DNA intermediates that would otherwise have a propensity to degenerate into toxic DNA lesions when unprotected (Chen et al., 2009; Pebernard et al., 2006; Torres-Rosell et al., 2005). The repair of RNA-DNA hybrids –or R-loops– represents a salient example of a process that could benefit from the DNA compaction activity of the Smc5/6 complex. Indeed, R-loop repair requires the reannealing of separated ssDNA molecules (Garcia-Muse and Aguilera, 2019) and DNA compaction by the Smc5/6 complex might enhance

formation of dsDNA by bringing complementary ssDNA together in a smaller effective volume, thus increasing the likelihood of their interaction. This prediction is consistent with our demonstration that the Smc5/6 complex can bind to RNA-DNA hybrids and ssDNA molecules, as well as genetic evidence connecting the complex to R-loop repair *in vivo* (Styles et al., 2016). We envision that the contribution of Smc5/6-mediated micro-compaction in DNA transactions will not be limited to R-loop repair and will likely involve several other DNA repair pathways.

It is remarkable that the behavior of the human Smc5/6 complex is so similar to that of the budding yeast Smc5/6 complex described in an accompanying paper by Gutierrez-Escribano et al. (2020). Both the yeast and human complexes associate tightly with nucleic acids, recognize supercoiled DNA structures and compact DNA with high efficiency. These are unusual biochemical activities to harbour in a single DNA repair enzyme. Likewise, both of our studies have identified Smc6 ATPase/CC neck as a likely interaction hub for other subunits of the Smc5/6 complex. Collectively, these similarities indicate that the absence of Nse2 subunit in the human complex does not negatively impact the biochemical activities reported herein, consistent with the fact that Nse2 is not essential for cell viability in metazoans (Kliszczak et al., 2012; Verver et al., 2016; Zheng et al., 2017). One possible difference we observed between the yeast and human complexes lies in the configuration of the SMC arms. The arms of the yeast complex seem to adopt a predominantly folded conformation, while our EM analysis suggests an extended conformation for human SMC arms. While this may appear as a discrepancy, recent structural studies indicate that SMC complexes can adopt distinct folded configurations (Burmann et al., 2019; Lee et al., 2020). Small differences in purification, imaging conditions or nucleotide status may favor one configuration over the other, thus explaining the different folded states of the yeast and human Smc5/6 complexes in our studies. In this context, the extended and folded CC arms of Smc5/6 complexes should be viewed as alternative physiological

states, not as mutually exclusive conformations. Altogether, the yeast and human studies entirely converge in their demonstrations that the Smc5/6 enzyme is a structure-specific DNA compacting machine, an exciting paradigm for SMC-mediated DNA repair reactions.

The clinical and biological implications of our findings are exemplified by the connection of the LIC syndrome to a plausible defect in the function or regulation of Smc6 ATPase head/neck region. Indeed, we have identified an interaction hub within Smc6 and shown that mutations in LIC syndrome patients map directly to the region of Nse3 that associates with the interaction hub in Smc6. The same region of Smc6 was proposed to play an important role as a potential gate for the control of DNA movement through SMC arms (reviewed in Palecek and Gruber, 2015). We infer from these observations that patients with LIC syndrome might harbor a defect in a Nse3-WH/Smc6 gate that controls transit of DNA into the Smc5/6 ring structure.

Limitations

Our DNA binding analyses indicate that the Smc5/6 complex associates tightly to nucleic acids using a standard –likely electrostatic– mode of binding. Our experiments did not address whether the Smc5/6 complex can bind DNA through topological entrapment (Kanno et al., 2015). Indeed, DNA association based exclusively on topological entrapment typically cannot be observed on linear DNA substrates (such as those used in Figure 5.5; Ocampo-Hafalla and Uhlmann, 2011), and additional experiments will be required to clarify this question. Future work should also focus on assessing the capacity of the Smc5/6 complex to induce DNA micro-compaction during DNA repair reactions. Our study provided a compelling demonstration of the capacity of the Smc5/6 complex to compact DNA *in vitro*, but *in vivo* visualization of this mechanism would further reinforce the paradigm we propose and provide a system to assess how ancillary effectors of the Smc5/6 complex (such as Nse2 and the elusive human homolog of yeast Nse5) might impact DNA repair reactions promoted by the core

Smc5/6 complex.

Acknowledgments

We thank members of the D'Amours laboratory for their comments on the manuscript and Peter Stirling (UBC) for sharing *nse1/3/4* mutant strains. This work was supported by grants from CIHR to DD (FDN–167265) and NSERC to JMP (DG RGPIN-2015-05776). Work at NU was supported by the NIH grants R01-GM105847, U54-CA193419 (CR-PS-OC) and a subcontract to grant U54-DK107980 (4D Nucleome). DD is supported by a Canada Research Chair in Chromatin Dynamics & Genome Architecture. DS & AS were supported by post-doctoral fellowships from the FRQS (#29086 & #33382) and NSERC (#487769-2016), respectively.

Author contributions

DS, RK, AS, JMP, JFM & DD conceived and designed the experiments; DS performed the *in vivo* fusion and XL-MS experiments; DS, MS & GC performed purifications, DNA-binding & enzymatic assays; GC carried out the analytical ultracentrifugation & GraFix procedures; AS & JMP performed EM imaging; RK carried out magnetic tweezers experiments; DS, SR & CM created yeast strains, assessed their viability & Smc6 abundance; DS, GC, RK, AS, JMP, JFM & DD analyzed the data; DS, RK, SR & AS prepared figures; DS & DD wrote the paper.

Declaration of Interests

The authors declare no competing interests.

STAR Methods

Resource availability

Lead contact: Requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Damien D'Amours (damien.damours@uottawa.ca).

Materials availability: Plasmids and yeast strains generated in this study are available on request. MTA may be required to share materials, in accordance with the University of Ottawa policy on inter-institutional transfer of research materials.

Data and code availability: The datasets generated during this study are available at Mendeley [http://dx.doi.org/10.17632/35z5bfkpv3.1].

Experimental model and subject details

Yeast strains. All yeast strains used in this study are derivative of K699/K700 (Table S2). Gene deletion strains were generated by PCR amplification as previously described (Longtine et al., 1998). Yeast growth conditions, media composition and procedures for genetic analysis were as published before (Robellet et al., 2015). For DNA-damaging experiments, yeast were dropped on solid medium containing different concentrations of MMS, 4-NQO and HU at 23 °C and 37 °C. Briefly, 5-fold dilution series of fusion yeast strains (first spot corresponds to a culture at OD_{600} of 0.2) were spotted on solid YPD (Yeast extract, Peptone, 2% Glucose) and grown in temperature-controlled incubators for 48-72 hours before scanning the plate.

For protein overexpression in fermentor, yeast strains were grown at 32 °C in YEP with 2% lactic acid and 3% glycerol as carbon source. When OD_{600} reached 0.5-0.6, protein overexpression was induced at 23 °C by addition of galactose to a final concentration of 1.5%. For protein

overexpression in incubator/shaker, yeast strains were grown at 30 °C/250 rpm in YEP with 2% raffinose as carbon source. When OD_{600} reached 0.7-0.9, protein overexpression was induced at 18 °C/250 rpm by addition of galactose to a final concentration of 2%. At the end of the experiment, yeast cells were harvested, washed with cold water and frozen by immersion in liquid nitrogen.

Method Details

Yeast fusions. Endogenous sequences (including *SMC6* and *NSE4* promotors) were amplified by PCR from strain D4107 (wild-type, haploid). Primers were designed with overhanging sequences to facilitate DNA fragment fusion by PCR and allow addition of flexible linker-encoding sequences between genes (while preventing the addition of amino-acids due to the presence of restriction sites). The same linker sequences as those used for fusion of human genes were used in the yeast fusion constructs (see plasmids p1409 and p1410 below). All PCR products were purified by gel extraction (Qiagen) and then a mix containing Herculase II buffer, dNTPs, PFU Herculase II polymerase, and 1:1 DNA ratio *Linker-*P_{NSE4}*-NSE4* or *Linker-*P_{SMC6}*-SMC6* was incubated for two PCR cycles. Then, primers were added to extend the joined-fragments product. A similar strategy was followed to fuse P_{NSE4}*-NSE4-Linker* to *SMC5*. Final constructs were cloned into the YIplac211 integrative plasmid using XmaI and SalI restriction enzymes (New England Biolabs).

smc6Δ, *nse4Δ* and *smc5Δ* were generated in diploid strains due to the fact that subunits of the Smc5/6 complex are essential. 4µg of the *URA3*-integrative plasmids YIplac211-P_{SMC6}*-SMC6-Linker-NSE4* and YIplac211-P_{NSE4}*-NSE4-Linker-SMC5* were linearized and transformed in *SMC6/smc6Δ NSE4/nse4Δ* and *SMC5/smc5Δ NSE4/nse4Δ* strains, respectively. Positive diploid strains expressing *SMC6-Linker-NSE4* and *NSE4-Linker-SMC5* were sporulated and dissected using a Nikon 50i microscope equipped with a tetrad micro-manipulator. Finally, spores carrying the double deletion

and the fusion protein were screened by PCR to confirm the genotype.

Purification of Smc5/6 dimer. Human Smc5 and Smc6 subunits were codon-optimized for expression in yeast and subcloned in a 2μ-derived plasmid under the control of the *GAL10-1* promoter sequence, as previously done for condensin subunits (St-Pierre et al., 2009). The Smc5/6 dimer was purified from ≥14L cultures of yeast grown to an OD_{600} of 1.0 – 1.5. Yeast overexpressing human Smc5/6 dimer were lysed in a freezer mill and resuspended in buffer A (500 mM NaCl, 100 mM Tris-HCl pH 8.0 and 10% glycerol) supplemented with 20 mM imidazole, 0.5% Triton X-100, 2 mM β-mercaptoethanol (βME) and protease inhibitors (E64, PEPA, AESBF). Lysate was centrifuged at 12,000rpm for 30 minutes at 4 °C. Next, Ni-NTA resin was added to the supernatant and incubated for one hour. Once binding was complete, resin was washed with 10CV of buffer A and then with 2CV of buffer ATP (buffer A supplemented with 2 mM ATP and 25 mM KCl). Protein was finally eluted in buffer A supplemented with 500 mM imidazole. Ni-NTA elution was then passed through a column loaded with avidin-agarose beads. Next, the avidin flow through material was loaded into a StrepTrap HP 5 mL column and washed with 10 CV of buffer A supplemented with 1 mM DTT and 0.5% Triton X-100 and then, it was followed by a wash with buffer A supplemented with 2 mM βME until UV was zero. Finally, proteins were eluted in buffer A supplemented with 2 mM βME and 20 mM desthiobiotin. Elutions containing the dimer were dialyzed in buffer A, concentrated, quantified, snap frozen and stored at -80 °C.

Purification of the Smc5/6 core complex. Subunits of the human Smc5/6 complex were codon-optimized for expression in yeast and subcloned in 2μ-derived plasmids under the control of the *GAL10-1* promoter sequence (see Fig. 5.1B for a description of genes contained in overexpression

plasmids and the tagging/fusion strategies). To create subunit fusions, we inserted the DNA sequence encoding GGGGGPRENLYFQGPRENLYFQGASENLYFQGGGGGGGG ASENLYFQGEAG and GGGGSGGGSGGGGTRARENLYFQGASENLYFQGELENLYFQGAS linkers between the coding sequences of Nse4a and Smc5 (plasmid p1409) and Smc6 and Nse4a (plasmid p1410), respectively. The Smc5/6 core complex was typically purified from 70L cultures of yeast grown to an OD_{600} of 1.0 – 1.5. Yeast pellets overexpressing the human Smc5/6 complex (Nse4a-L-Smc5 or Smc6-L-Nse4a fusions) were lysed in a freezer mill and resuspended in buffer N (50 mM K_2HPO_4 / KH_2PO_4 pH8, 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol, 0.5 % triton X-100, 2 mM βME) supplemented with 20 mM imidazole and protease inhibitors (E64, PEPA, AESBF). Lysate was centrifuged at 12000 rpm for 30 min at 4 °C and then Ni-NTA resin was loaded to the supernatant and incubated for one hour. Resin was washed with 10 column volumes (CV) of buffer N supplemented with 60 mM imidazole. Complex was eluted with buffer SB (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol, 0.5% tween 20, 2mM βME) supplemented with 500 mM imidazole. Flow through of Ni-NTA purification was loaded with Ni-NTA resin for a second round of purification. Next, elution fractions were loaded into a StrepTrap HP 5mL column and washed with 10 CV of buffer SB supplemented with 0.5% Triton X-100. Proteins were eluted with 5 CV of buffer GB (25 mM K_2HPO_4/KH_2PO_4 pH8, 500 mM NaCl, 10% glycerol and 2 mM βME) supplemented with 20 mM desthiobiotin. Next, the elution was loaded in a GSTrap 5 mL column, washed with 10 CV of buffer GB and the Smc5/6 complex was eluted with 5 CV of buffer GEB (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol and 2 mM βME supplemented with 10 mM of Glutathione). Linker, poly-histidine, Strep-tag II and GST tags were cleaved by an overnight digestion with 1 mg of TEV protease per 4 mg/mL of fusion protein. Digestion was carried out in GEB buffer supplemented with 1 mM DTT. Digestion product was loaded into a Superose 6 10/300 size exclusion chromatography column in GF buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol and 2 mM

βME) in order to remove the cleaved tags, digested linker and TEV. Elution fractions containing highly purified and stoichiometric complex were concentrated, quantified, snap frozen and stored at -80 °C. Although purification experiments shown herein were conducted with Smc5/6 complexes derived from the Nse4a-linker-Smc5 fusion variant, we noticed that complexes derived from the Smc6-linker-Nse4a fusion could also be efficiently purified using the same procedure.

Density gradients. Human Smc5/6 complex (80 pmoles) was applied on a continuous 5-20% sucrose density gradient (50 mM Tris-HCl pH 8, 500 mM NaCl, 0.05 mM EDTA, 6.25 mM $MgCl_2$, 0.5 mM βME). The gradient was generated in a BioComp gradient station. Ultracentrifugation was performed at 4 °C during 22 hours at 32,000 rpm in a Sorvall WX100 centrifuge (SW 41 Ti rotor, Beckman Coulter). 450 µL fractions were collected manually by pipetting from the top of the tube to the bottom and visualized by silver staining using SilverQuest Kit (Invitrogen).

A modified GraFix protocol was followed when complex stabilization was required for structural analysis (Kastner et al., 2008; Stark, 2010). Briefly, the heavy fraction (20% sucrose) of a 5-20% sucrose gradient (50 mM HEPES pH8, 500 mM NaCl, 0.05 mM EDTA, 6.25 mM $MgCl_2$, 0.5 mM βME) was supplemented with 0.15% glutaraldehyde. The gradient was generated in a BioComp gradient station. The samples were ultracentrifuged and fractionated as above and immediately quenched by addition of 450 µL of buffer Q (100 mM Tris-HCl pH 8, 500 mM glycine, 300 mM NaCl and 2% glycerol). Samples were visualized by silver staining.

Negative stain electron microscopy and image analysis. Carbon-coated copper grids (Electron Microscopy Sciences, USA) were negatively glow-discharged (Agar Scientific, USA) before adsorbing 5 µl of sample, then staining with 5 µl of freshly prepared 1.5% uranyl formate (Electron Microscopy

Sciences, USA). Samples were imaged at room temperature using a FEI Tecnai T12 (Eindhoven, The Netherlands) Transmission Electron Microscope equipped with a LaB6 filament and operated at 120 kV. Images were collected at defocus between 0.5–2 µm on a FEI Eagle 4k x 4k CCD camera at magnifications of ~67,000 x (pixel size 1.64 Å) or ~110,000 x (pixel size 0.99 Å). Particles were selected manually using EMAN2 (Tang et al., 2007) and extracted in 518 Å x 518 Å boxes with RELION 3.0 (Scheres, 2012). 2D averages were generated using 44 Smc5/6 complex pre-GraFix particles, 68 Smc5/6 complex post-GraFix particles, and 167 Smc5/6 dimer post-GraFix particles. Contrast transfer function correction was not implemented during image processing. For 3D image analysis, 945 post-GraFix Smc5/6 complex particles were imaged at ~42,000x (pixel size 2.59 Å) and aligned using the CL2D algorithm (Sorzano et al., 2010) from Xmipp (Sorzano et al., 2004). An initial 3D model was generated from 2D classes using RANSAC (Vargas et al., 2014) and the volume was refined with single particles in RELION 3.0 (Scheres, 2012) after low-pass filtering to 60 Å. Default refinement parameters were used without masking nor imposing symmetry. The final 3D EM map was visualized using UCSF Chimera (Pettersen et al., 2004).

DNA binding and ATPase assays. The DNA binding activity of human Smc5/6 complex was determined by electrophoretic mobility shift assay (EMSA) according to a modified protocol previously used in our laboratory (Roy and D'Amours, 2011; Roy et al., 2011). DNA substrates were ϕX174 (ssDNA substrate; 5386 bp) and EcoRI-digested pBluescript II KS (dsDNA substrate; 2961 bp). Briefly, protein-DNA binding assays were carried out in 25 µL of reaction buffer (10 mM Hepes pH 7.5, 160 mM NaCl, 7 mM $MgCl_2$, 20% glycerol and 2 mM βME) containing 50 ng of DNA substrate (either ss or dsDNA) and various molar excess amounts of Smc5/6 dimer or core complex. After incubation at 30 °C for 30 min, the reactions were terminated by addition of an equal volume of 1.6%

low melting point agarose containing 1 µL of loading buffer (0.6% glycerol, 0.005% bromophenol blue, 0.005% xylene cyanol). The mixture was then loaded on a 0.8% TAE-agarose gel and the DNA was resolved by electrophoresis for 16 hours at 4 °C. DNA was stained in agarose gels with SYBR Gold (ssDNA) or SYBR Green I (dsDNA) reagents and the resulting fluorescent signal was imaged using a Thyphoon FLA 9500 scanner. Free DNA was quantified using ImageJ software. For each condition, unbound DNA was quantified and plotted as a percentage of the total DNA loaded in the no-protein lane. Curve fitting of affinity constants and Hill coefficients were determined using GraphPad Prism 7 (GraphPad Software Inc). ATPase assays were performed as previously described with minor modifications (Roy et al., 2011). In this assay, the Smc5/6 ATPase showed a nucleotide hydrolysis rate of 0.049 ± 0.009 mol/sec/mol of protein, similar in range to the activity reported for condensin (Kimura and Hirano, 2000), and vastly superior to the activity of the individual Smc5 monomer (0.0019 ± 0.0005 mol of ATP per mol of Smc5 per second, as measured by (Roy et al., 2011)).

Magnetic tweezers experiments. Linear DNA fragments used in single-molecule supercoiling relaxation experiments were derived from the plasmid pNG1175 (9702 bp), a slightly modified version of pFOS-1 (9691 bp, New England Biolabs) (Keenholtz et al., 2017). pNG1175 was linearized by cutting at nearby SpeI and ApaI restriction sites; the resulting linear molecule was ligated to ≈900 bp PCR products carrying either biotinylated or digoxigenin-labeled nucleotides, prepared with SpeI and ApaI-compatible ends, respectively. The resulting linear constructs were 11.4 kb in length, with roughly 900-bp of biotin- and digoxigenin-labeled DNA at their ends, allowing multiple tethering of the ends to streptavidin- or anti-digoxigenin-coated surfaces. The multiple tethers constrain the two DNA strands sufficiently that they may be supercoiled by rotation of the magnetic particle.

Flow cells were assembled for each experiment and contained 2.8 µm streptavidin coated

paramagnetic beads (Invitrogen Dynabeads, M-270) tethered to the surface of an anti-digoxigenin coated glass coverslip via a linear pNG1175 DNA molecule with biotinylated and digoxigenin-labeled ends (Fig. 5.6A). Flow cell contents were viewed with a bright field microscope and a 100× 1.3NA oil immersion objective (Olympus). Translation in the z direction of a permanent magnet under the objective stage controlled the force on the bead while 360° rotations of the magnet controlled the linking number of the tethered DNA molecule. Bead position in three dimensions was tracked with custom lab-written software (LabView, National Instruments), which uses an untethered bead nonspecifically bound to the glass surface as a reference point. Position fluctuations in the x-y plane were used to calibrate the force on the tethered beads while changes in the z direction relative to the reference bead were used to measure the tether extension (Skoko et al., 2004). Experiments recorded DNA extension for approximately 1000 sec, at approximately 100 measurements per second.

Single-molecule experiments to study DNA compaction were carried out in an assay buffer containing 10 mM HEPES pH 7.5 and 50 mM potassium glutamate at 25 °C. An additional 1 mM nucleotide (ATP, ADP or ATPγS) along with 1 mM $MgCl_2$ was included as noted. The Smc5/6 core complex was added to a 200 µL mixture of assay buffer for a final concentration of 5 nM and immediately added to flow cells with tethers initially held at 4 pN force, with force reduced to 0.30 pN (or other forces as indicated for force-titration experiments) following addition of enzyme solution to the flow cell.

There are variations in tether length due to adhering of varied amounts of the 900 bp (300 nm) labeled ends, as well as due to random variation in the bead sizes, position on the bead of DNA tethering, and bead optical properties, which affect the precise location of the focal plane that determines the inferred position of the beads in the vertical direction. These effects lead to variation

of initial length in the few hundred nanometer range as observed.

Single-molecule supercoil/plectoneme capture experiments were carried out following the same method except for the use of a buffer containing 100 mM potassium glutamate. The effect of increased potassium glutamate on single-molecule experiments could act at two levels: stimulation of an otherwise silent activity in the Smc5/6 complex and/or suppression of another activity in the complex. Potassium glutamate is known to stabilize many protein-DNA complexes (Cheng et al., 2016 and reference cited therein). Glutamate also has the ability to relieve inhibition induced by some salts in enzymatic reactions (Griep and McHenry, 1989). It is conceivable that higher concentrations of potassium glutamate in single-molecule experiments could stabilize Smc5/6 complex binding to plectonemic DNA or stimulate a dormant activity in the complex, but more work is required to define the detailed effects of buffers/salts on Smc5/6 activity.

Protein crosslinking and mass spectrometry. Proteins were purified as described above. A total of 50 µg of protein in buffer C (50 mM HEPES pH8, 500 mM NaCl, 5% glycerol) was crosslinked at room temperature for 5 minutes with 0.3 mM DSS (Thermo Fisher) or with DMTMM (Sigma Aldrich). Note that DSS is a Lys-Lys crosslinker whereas DMTMM is a Lys-Asp or Lys-Glu crosslinker. Crosslinking reactions were then quenched by addition of 10 µL of 1 M Tris-HCl pH 8. Samples were then analyzed by mass spectrometry at the Proteomics Core Facility in IRIC (Montreal, Canada) by scientists with experience in the XL-MS procedure (Courcelles et al., 2017). To ensure our XL-MS analysis does not lead to false-positive identifications, all procedures were performed with enzyme preparations that are active in biochemical assays reported herein. Furthermore, MS spectra identified in our XL analysis were inspected manually and their identity re-confirmed by direct observation to ensure they truly represent crosslinked peptides, as previously described (Courcelles et al., 2017). Finally, XL

reactions were not conducted to full saturation to avoid capture of spurious interactions and formation of unspecific XLs. Detailed circular maps of inter-subunit and intra-protein crosslinks were generated using Circos software (Krzywinski et al., 2009).

Immunoblot analysis. Cell lysates were prepared from exponential cultures of yeast grown at 23 °C or 37 °C using the TCA glass-bead method (Foiani et al., 1994). Lysates were subsequently resolved by SDS-PAGE and processed for immunoblot analysis using an anti-HA antibody 12CA5 (1:2500 dilution; MilliporeSigma) and an anti-mouse IgG antibody (1:5,000 dilution; Cytiva), as described by St-Pierre et al. (2009). A 12CA5 cross-reacting band was used as a loading control, as previously described (Gallego et al., 1997). Band intensity on immunoblots was measured using Image J software (NIH, USA).

Quantification and statistical analysis

Data are presented as means ± SEM. All statistical analyzes were performed using GraphPad Prism 7 (GraphPad Software Inc) and statistical significance threshold was set at *p*-value = 0.05. Where indicated in figure legends, we performed an analysis of variance (ANOVA) followed by a Dunnett's *post hoc* test.

References

Alt, A., Dang, H.Q., Wells, O.S., Polo, L.M., Smith, M.A., McGregor, G.A., Welte, T., Lehmann, A.R., Pearl, L.H., Murray, J.M.*, et al.* (2017). Specialized interfaces of Smc5/6 control hinge stability and DNA association. Nat Commun *8*, 14011.

Andrews, E.A., Palecek, J., Sergeant, J., Taylor, E., Lehmann, A.R., and Watts, F.Z. (2005). Nse2, a component of the Smc5-6 complex, is a SUMO ligase required for the response to DNA damage. Mol Cell Biol *25*, 185-196.

Aragon, L. (2018). The Smc5/6 Complex: New and Old Functions of the Enigmatic Long-Distance Relative. Annu Rev Genet *52*, 89-107.

Baxter, J., Oliver, A.W., and Schalbetter, S.A. (2019). Are SMC Complexes Loop Extruding Factors? Linking Theory With Fact. Bioessays *41*, e1800182.

Behlke-Steinert, S., Touat-Todeschini, L., Skoufias, D.A., and Margolis, R.L. (2009). SMC5 and MMS21 are required for chromosome cohesion and mitotic progression. Cell Cycle *8*, 2211-2218.

Bitard-Feildel, T., Lamiable, A., Mornon, J.P., and Callebaut, I. (2018). Order in Disorder as Observed by the "Hydrophobic Cluster Analysis" of Protein Sequences. Proteomics *18*, e1800054.

Burmann, F., Basfeld, A., Vazquez Nunez, R., Diebold-Durand, M.L., Wilhelm, L., and Gruber, S. (2017). Tuned SMC Arms Drive Chromosomal Loading of Prokaryotic Condensin. Mol Cell *65*, 861-872 e869.

Burmann, F., Lee, B.G., Than, T., Sinn, L., O'Reilly, F.J., Yatskevich, S., Rappsilber, J., Hu, B., Nasmyth, K., and Lowe, J. (2019). A folded conformation of MukBEF and cohesin. Nat Struct Mol Biol *26*, 227-236.

Chen, Y.H., Choi, K., Szakal, B., Arenz, J., Duan, X., Ye, H., Branzei, D., and Zhao, X. (2009). Interplay between the Smc5/6 complex and the Mph1 helicase in recombinational repair. Proc Natl Acad Sci U S A *106*, 21252-21257.

Cheng, X., Guinn, E.J., Buechel, E., Wong, R., Sengupta, R., Shkel, I.A., and Record, M.T., Jr. (2016). Basis of Protein Stabilization by K Glutamate: Unfavorable Interactions with Carbon, Oxygen Groups. Biophys J *111*, 1854-1865.

Courcelles, M., Coulombe-Huntington, J., Cossette, E., Gingras, A.C., Thibault, P., and Tyers, M. (2017). CLMSVault: A Software Suite for Protein Cross-Linking Mass-Spectrometry Data Analysis and Visualization. J Proteome Res *16*, 2645-2652.

de la Rosa-Trevin, J.M., Quintana, A., Del Cano, L., Zaldivar, A., Foche, I., Gutierrez, J., Gomez-Blanco, J., Burguet-Castell, J., Cuenca-Alba, J., Abrishami, V.*, et al.* (2016). Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. J Struct Biol *195*, 93-99.

Doyle, J.M., Gao, J., Wang, J., Yang, M., and Potts, P.R. (2010). MAGE-RING protein complexes comprise a family of E3 ubiquitin ligases. Mol Cell *39*, 963-974.

Eeftens, J.M., Bisht, S., Kerssemakers, J., Kschonsak, M., Haering, C.H., and Dekker, C. (2017). Real-time detection of condensin-driven DNA compaction reveals a multistep binding mechanism. Embo j *36*, 3448-3457.

Foiani, M., Marini, F., Gamba, D., Lucchini, G., and Plevani, P. (1994) The B subunit of the DNA polymerase alpha-primase complex in Saccharomyces cerevisiae executes an essential function at the initial stage of DNA replication. Mol Cell Biol 14, 923–933.

Gallego, C., Garí, E., Colomina, N., Herrero, E., and Aldea, M. (1997) The Cln3 cyclin is down-regulated by translational repression and degradation during the G1 arrest caused by nitrogen deprivation in budding yeast. EMBO J 16, 7196–7206.

Garcia-Muse, T., and Aguilera, A. (2019). R Loops: From Physiological to Pathological Roles. Cell *179*, 604-618.

Griep, M.A., and McHenry, C.S. (1989). Glutamate overcomes the salt inhibition of DNA polymerase III holoenzyme. J Biol Chem *264*, 11294-11301.

Gruber, S., Arumugam, P., Katou, Y., Kuglitsch, D., Helmhart, W., Shirahige, K., and Nasmyth, K. (2006). Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. Cell *127*, 523-537.

Hudson, J.J., Bednarova, K., Kozakova, L., Liao, C., Guerineau, M., Colnaghi, R., Vidot, S., Marek, J., Bathula, S.R., Lehmann, A.R.*, et al.* (2011). Interactions between the Nse3 and Nse4 components of the SMC5-6 complex identify evolutionarily conserved interactions between MAGE and EID Families. PLoS One *6*, e17270.

Irmisch, A., Ampatzidou, E., Mizuno, K., O'Connell, M.J., and Murray, J.M. (2009). Smc5/6 maintains stalled replication forks in a recombination-competent conformation. Embo J *28*, 144-155.

Jeppsson, K., Carlborg, K.K., Nakato, R., Berta, D.G., Lilienthal, I., Kanno, T., Lindqvist, A., Brink, M.C., Dantuma, N.P., Katou, Y.*, et al.* (2014). The chromosomal association of the Smc5/6 complex depends on cohesion and predicts the level of sister chromatid entanglement. PLoS Genet *10*, e1004680.

Kanno, T., Berta, D.G., and Sjogren, C. (2015). The Smc5/6 Complex Is an ATP-Dependent Intermolecular DNA Linker. Cell Rep *12*, 1471-1482.

Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E.*, et al.* (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. Nat Methods *5*, 53-55.

Keenholtz, R.A., Dhanaraman, T., Palou, R., Yu, J., D'Amours, D., and Marko, J.F. (2017). Oligomerization and ATP stimulate condensin-mediated DNA compaction. Sci Rep *7*, 14279.

Kegel, A., Betts-Lindroos, H., Kanno, T., Jeppsson, K., Strom, L., Katou, Y., Itoh, T., Shirahige, K., and Sjogren, C. (2011). Chromosome length influences replication-induced topological stress. Nature *471*, 392-396.

Kimura, K., and Hirano, T. (2000). Dual roles of the 11S regulatory subcomplex in condensin functions. Proc Natl Acad Sci U S A *97*, 11972-11977.

Kliszczak, M., Stephan, A.K., Flanagan, A.M., and Morrison, C.G. (2012). SUMO ligase activity of vertebrate Mms21/Nse2 is required for efficient DNA repair but not for Smc5/6 complex stability. DNA Repair (Amst) *11*, 799-810.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res *19*, 1639-1645.

Lafuente-Barquero, J., Luke-Glaser, S., Graf, M., Silva, S., Gomez-Gonzalez, B., Lockhart, A., Lisby, M., Aguilera, A., and Luke, B. (2017). The Smc5/6 complex regulates the yeast Mph1 helicase at RNA-DNA hybrid-mediated DNA damage. PLoS Genet *13*, e1007136.

Lee, B.G., Merkel, F., Allegretti, M., Hassler, M., Cawood, C., Lecomte, L., O'Reilly, F.J., Sinn, L.R., Gutierrez-Escribano, P., Kschonsak, M.*, et al.* (2020). Cryo-EM structures of holo condensin reveal a subunit flip-flop mechanism. Nat Struct Mol Biol *27*, 743-751.

Longtine, M.S., Fares, H., and Pringle, J.R. (1998). Role of the yeast Gin4p protein kinase in septin assembly and the relationship between septin assembly and septin function. J Cell Biol *143*, 719-736.

Marko, J.F., De Los Rios, P., Barducci, A., and Gruber, S. (2019). DNA-segment-capture model for loop extrusion by structural maintenance of chromosome (SMC) protein complexes. Nucleic Acids Res *47*, 6956-6972.

Newman, J.A., Cooper, C.D., Roos, A.K., Aitkenhead, H., Oppermann, U.C., Cho, H.J., Osman, R., and Gileadi, O. (2016). Structures of Two Melanoma-Associated Antigens Suggest Allosteric Regulation of Effector Binding. PLoS One *11*, e0148762.

Ocampo-Hafalla, M.T., and Uhlmann, F. (2011). Cohesin loading and sliding. J Cell Sci *124*, 685-691.

Palecek, J.J., and Gruber, S. (2015). Kite Proteins: a Superfamily of SMC/Kleisin Partners Conserved Across Bacteria, Archaea, and Eukaryotes. Structure *23*, 2183-2190.

Pebernard, S., Wohlschlegel, J., McDonald, W.H., Yates, J.R., 3rd, and Boddy, M.N. (2006). The Nse5-Nse6 dimer mediates DNA repair roles of the Smc5-Smc6 complex. Mol Cell Biol *26*, 1617-1630.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Potts, P.R., Porteus, M.H., and Yu, H. (2006). Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks. EMBO J *25*, 3377-3388.

Robellet, X., Thattikota, Y., Wang, F., Wee, T.L., Pascariu, M., Shankar, S., Bonneil, E., Brown, C.M., and D'Amours, D. (2015). A high-sensitivity phospho-switch triggered by Cdk1 governs chromosome morphogenesis during cell division. Genes Dev *29*, 426-439.

Roy, M.A., and D'Amours, D. (2011). DNA-binding properties of Smc6, a core component of the Smc5-6 DNA repair complex. Biochem Biophys Res Commun *416*, 80-85.

Roy, M.A., Dhanaraman, T., and D'Amours, D. (2015). The Smc5-Smc6 heterodimer associates with DNA through several independent binding domains. Sci Rep *5*, 9797.

Roy, M.A., Siddiqui, N., and D'Amours, D. (2011). Dynamic and selective DNA-binding activity of Smc5, a core component of the Smc5-Smc6 complex. Cell Cycle *10*, 690-700.

Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol *180*, 519-530.

Skoko, D., Wong, B., Johnson, R.C., and Marko, J.F. (2004). Micromechanical analysis of the binding of DNA-bending proteins HMGB1, NHP6A, and HU reveals their ability to form highly stable DNA-protein complexes. Biochemistry *43*, 13867-13874.

Skoko, D., Yoo, D., Bai, H., Schnurr, B., Yan, J., McLeod, S.M., Marko, J.F., and Johnson, R.C. (2006). Mechanism of chromosome compaction and looping by the Escherichia coli nucleoid protein Fis. J Mol Biol *364*, 777-798.

Sorzano, C.O., Bilbao-Castro, J.R., Shkolnisky, Y., Alcorlo, M., Melero, R., Caffarena-Fernandez, G., Li, M., Xu, G., Marabini, R., and Carazo, J.M. (2010). A clustering approach to multireference alignment of single-particle projections in electron microscopy. J Struct Biol *171*, 197-206.

Sorzano, C.O., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H., Carazo, J.M., and Pascual-Montano, A. (2004). XMIPP: a new generation of an open-source image processing package for electron microscopy. J Struct Biol *148*, 194-204.

St-Pierre, J., Douziech, M., Bazile, F., Pascariu, M., Bonneil, E., Sauve, V., Ratsima, H., and D'Amours, D. (2009). Polo kinase regulates mitotic chromosome condensation by hyperactivation of condensin DNA supercoiling activity. Mol Cell *34*, 416-426.

Stark, H. (2010). GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. Methods Enzymol *481*, 109-126.

Strick, T.R., Allemand, J.F., Bensimon, D., Bensimon, A., and Croquette, V. (1996). The elasticity of a single supercoiled DNA molecule. Science *271*, 1835-1837.

Strick, T.R., Kawaguchi, T., and Hirano, T. (2004). Real-time detection of single-molecule DNA compaction by condensin I. Curr Biol *14*, 874-880.

Styles, E.B., Founk, K.J., Zamparo, L.A., Sing, T.L., Altintas, D., Ribeyre, C., Ribaud, V., Rougemont, J., Mayhew, D., Costanzo, M.*, et al.* (2016). Exploring Quantitative Yeast Phenomics with Single-Cell Analysis of DNA Damage Foci. Cell Syst *3*, 264-277 e210.

Sun, M., Nishino, T., and Marko, J.F. (2013). The SMC1-SMC3 cohesin heterodimer structures DNA through supercoiling-dependent loop formation. Nucleic Acids Res *41*, 6149-6160.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol *157*, 38-46.

Torres-Rosell, J., Machin, F., Farmer, S., Jarmuz, A., Eydmann, T., Dalgaard, J.Z., and Aragon, L. (2005). SMC5 and SMC6 genes are required for the segregation of repetitive chromosome regions. Nat Cell Biol *7*, 412-419.

van der Crabben, S.N., Hennus, M.P., McGregor, G.A., Ritter, D.I., Nagamani, S.C., Wells, O.S., Harakalova, M., Chinn, I.K., Alt, A., Vondrova, L.*, et al.* (2016). Destabilized SMC5/6 complex leads to chromosome breakage syndrome with severe lung disease. J Clin Invest *126*, 2881-2892.

Vargas, J., Alvarez-Cabrera, A.L., Marabini, R., Carazo, J.M., and Sorzano, C.O. (2014). Efficient initial volume determination from electron microscopy images of single particles. Bioinformatics *30*, 2891-2898.

Verver, D.E., Zheng, Y., Speijer, D., Hoebe, R., Dekker, H.L., Repping, S., Stap, J., and Hamer, G. (2016). Non-SMC Element 2 (NSMCE2) of the SMC5/6 Complex Helps to Resolve Topological Stress. Int J Mol Sci *17*, 1782.

Yatskevich, S., Rhodes, J., and Nasmyth, K. (2019). Organization of Chromosomal DNA by SMC Complexes. Annu Rev Genet *53*, 445-482.

Zabrady, K., Adamus, M., Vondrova, L., Liao, C., Skoupilova, H., Novakova, M., Jurcisinova, L., Alt, A., Oliver, A.W., Lehmann, A.R., *et al.* (2016). Chromatin association of the SMC5/6 complex is dependent on binding of its NSE3 subunit to DNA. Nucleic Acids Res *44*, 1064-1079.

Zheng, Y., Jongejan, A., Mulder, C.L., Mastenbroek, S., Repping, S., Wang, Y., Li, J., and Hamer, G. (2017). Trivial role for NSMCE2 during in vitro proliferation and differentiation of male germline stem cells. Reproduction *154*, 181-195.

Table 5.S1. Crosslinks identified in the subunits of the Smc5/6 complex*, Related to Figure 5.3.

Dimer:	**Inter-subunit crosslinks (9)**				
Protein 1	***Peptide 1***	***Residue 1***	***Protein 2***	***Peptide 2***	***Residue 2***
Smc6	**K**MEEQQVR	310	Smc5	N**K**LESDYMAASSQLR	783
Smc5	ATDI**K**EASQK	332	Smc6	RG**K**R	815
Smc5	**K**LKEGQIPVTCR	298	Smc6	SL**K**IEAENK	773
Smc6	QL**K**ELK	467	Smc5	Y**K**QDVER	247
Smc6	EL**K**DSK	470	Smc5	**M**ATPSK	1
Smc6	**K**MEEQQVR	310	Smc5	NKLES**D**YMAASSQLR	787
Smc5	**E**IDVAK	139	Smc6	R**K**EENFSSPK	5
Smc5	IQ**D**EK	412	Smc6	**K**NISEIR	719
Smc5	**E**RETLEK	427	Smc6	IEVE**K**SASILDK	862
Dimer:	**Intra-subunit crosslinks (51)**				
Protein 1	***Peptide 1***	***Residue 1***	***Protein 2***	***Peptide 2***	***Residue 2***
Smc5	DPSSEVPS**K**R	31	Smc5	TSTPSPQPS**K**R	17
Smc5	DPSSEVPS**K**R	31	Smc5	**K**STTQK	154
Smc5	**K**VELDQYR	911	Smc5	QLETSC**K**EK	230
Smc5	L**K**EGQIPVTCR	300	Smc5	V**K**EEVR	293
Smc5	DRV**K**EEVR	293	Smc5	**K**LKEGQIPVTCR	298
Smc6	HYEE**K**QK	821	Smc6	**K**MEEQQVR	310
Smc6	HYEEKQ**K**EHLDTLNK	823	Smc6	**K**MEEQQVR	310
Smc6	F**K**INQLSELADPLKDELNLADSEVDN	786	Smc6	YKDIQD**K**LEK	331
Smc6	FKINQLSELADPL**K**DELNLADSEVDN	798	Smc6	YKDIQD**K**LEK	331
Smc5	EASQ**K**CK	337	Smc5	Q**K**QDVIER	341
Smc5	A**K**LVTELTNLIK	748	Smc5	Q**K**QDVIER	341
Smc6	MVEEHMEQQ**K**ENMEHLK	763	Smc6	SLNEY**K**ALK	375
Smc6	DDEQLC**K**R	386	Smc6	**K**STDQSLEPER	393
Smc6	DS**K**TDR	473	Smc6	QL**K**ELK	467
Smc5	L**K**QIYTAEEK	610	Smc5	SSYAD**K**APSR	548
Smc5	L**K**QIYTAEEK	610	Smc5	YVV**K**TSFYSNK	622
Smc5	ALCEGEIID**K**R	424	Smc5	**K**LQAVDSGLIALR	665
Smc6	**K**ISWLK	409	Smc6	**K**NISEIR	719
Smc5	A**K**LVTELTNLIK	748	Smc5	I**K**EINVQK	740
Smc6	F**K**INQLSELADPLKDELNLADSEVDN	786	Smc6	IEAEN**K**YDAIK	779
Smc6	Q**K**IQAEHASHGDR	877	Smc6	SASILD**K**EINR	869
Smc5	**K**VELDQYR	911	Smc5	REEEIEQLTEEL**K**GK	908
Smc5	**E**IDVAK	139	Smc5	**K**TSTPSPQPSKR	7
Smc5	EI**D**VAK	141	Smc5	**K**TSTPSPQPSK	7
Smc5	**D**PSSEVPSK	23	Smc5	TSTPSPQPS**K**R	17

Smc5	DPSS**E**VPSK	27	Smc5	TSTPSPQPS**K**R	17
Smc5	**E**IDVAK	139	Smc5	KTSTPSPQPS**K**R	17
Smc5	DPSSEVPS**K**R	31	Smc5	**E**IDVAK	139
Smc6	GQTTLPFRPVTQ**E**EDDDQR	1085	Smc6	L**K**SATGSVVSTR	157
Smc6	GQTTLPFRPVTQE**E**DDDQR	1086	Smc6	L**K**SATGSVVSTR	157
Smc6	GQTTLPFRPVTQEE**D**DDQR	1087	Smc6	L**K**SATGSVVSTR	157
Smc6	GQTTLPFRPVTQEED**D**DQR	1088	Smc6	L**K**SATGSVVSTR	157
Smc5	IE**E**MENER	312	Smc5	L**K**EGQIPVTCR	300
Smc5	ATDI**K**EASQK	332	Smc5	IE**E**MENER	312
Smc6	LNEA**E**QK	322	Smc6	Y**K**DIQDKLEK	326
Smc6	APECMAL**K**ADVVAK	350	Smc6	IS**E**ETNAR	337
Smc6	IE**E**LKK	390	Smc6	KDDEQLC**K**R	386
Smc6	I**E**ELK	389	Smc6	**K**STDQSLEPER	393
Smc6	IE**E**LK	390	Smc6	**K**STDQSLEPER	393
Smc6	RIEEL**K**K	392	Smc6	ST**D**QSLEPER	396
Smc6	**K**ISWLK	409	Smc6	ST**D**QSLEPER	396
Smc6	**K**ISWLK	409	Smc6	STDQSL**E**PER	400
Smc5	G**K**KVELDQYR	910	Smc5	IQD**E**K	413
Smc6	D**K**EEHGKIKR	442	Smc6	**E**ELDVK	451
Smc6	DKEEHG**K**IKR	447	Smc6	**E**ELDVK	451
Smc5	IKEINVQ**K**AK	746	Smc5	LM**E**QDTCNLEEEER	722
Smc5	**K**ASTK	734	Smc5	LMEQDTCNL**E**EEER	729
Smc5	**K**ASTK	734	Smc5	LMEQDTCNLE**E**EER	730
Smc6	LT**E**LK	246	Smc6	SASILD**K**EINR	869
Smc6	IEV**E**K	861	Smc6	SASILD**K**EINR	869
Smc6	EKELEE**K**MSQAR	846	Smc6	**E**TYLDLDSK	900

Pentamer	**Inter-subunit crosslinks (39)**				
Protein 1	***Peptide 1***	***Residue 1***	***Protein 2***	***Peptide 2***	***Residue 2***
Smc5	**M**ATPSK	1	Smc6	RQCVE**K**EER	254
Smc5	DPSSEVPS**K**R	31	Nse4a	**K**VPVIQEER	238
Nse3	SQ**K**QLELK	83	Smc5	V**K**FR	976
Smc6	GSSL**K**GFVK	107	Nse3	HLIFGDP**K**K	220
Smc6	GSSL**K**GFVK	107	Nse3	LGVYPT**K**K	211
Nse1	**K**EAEQVLQK	144	Nse3	**K**HLIFGDPK	212
Nse1	**K**EAEQVLQK	144	Nse3	LGVYPT**K**K	211
Nse4a	A**K**QLR	149	Smc6	QL**K**ELKDSK	467
Smc6	NEGD**K**YK	206	Smc5	VGEFA**K**LSK	189
Nse3	**K**HLIFGDPK	212	Nse1	**K**KEAEQVLQK	143
Smc5	E**K**EK	221	Smc6	T**K**EQIHQGEER	234
Smc5	E**K**TEYLQK	232	Smc6	T**K**EQIHQGEER	234
Smc5	E**K**TEYLQK	232	Smc6	LTEL**K**R	248
Smc6	T**K**EQIHQGEER	234	Nse3	TNLETS**K**MK	260
Smc5	TEYLQ**K**MVQR	238	Smc6	T**K**EQIHQGEER	234
Nse4a	**K**VPVIQEER	238	Nse3	SQ**K**QLELK	83
Nse4a	**K**VPVIQEER	238	Smc6	NEGD**K**YK	206
Smc6	LTEL**K**R	248	Smc5	Y**K**QDVER	247
Smc5	DRV**K**EEVR	293	Smc6	HYEE**K**QK	821
Smc5	DRV**K**EEVR	293	Nse3	LQ**K**PR	4
Smc6	**K**MEEQQVR	310	Smc5	L**K**EGQIPVTCR	300
Smc6	**K**MEEQQVR	310	Smc5	N**K**LESDYMAASSQLR	783
Smc6	DKEEHG**K**IK	447	Smc5	DN**K**K	531
Smc5	ED**K**LR	455	Smc6	EELDV**K**HALSYNQR	456
Smc6	QL**K**ELK	467	Smc5	Y**K**QDVER	247
Smc6	AVMQSQ**K**PPK	628	Smc5	VNAVIAP**K**SSYADK	542
Smc5	ISS**K**LGSLK	714	Smc6	**K**STDQSLEPER	393
Smc6	HYEE**K**QK	821	Smc5	N**K**LESDYMAASSQLR	783

Smc5	DPSSEVPS**K**R	31	Nse4a	TF**E**ISEPVITPSQR	368
Nse3	DGFA**E**EAPSTSR	45	Smc6	**K**MEEQQVR	310
Nse3	DGFAE**E**APSTSR	46	Smc6	**K**MEEQQVR	310
Smc5	**E**IDVAK	139	Smc6	R**K**EENFSSPK	5
Nse1	**E**GEFTLHGR	163	Nse4a	**K**VPVIQEER	238
Nse1	EG**E**FTLHGR	165	Nse4a	**K**VPVIQEER	238
Nse1	**E**TYPDAVK	182	Nse4a	**K**VPVIQEER	238
Nse1	ETYP**D**AVK	186	Nse4a	**K**VPVIQEER	238
Smc6	**K**MEEQQVR	310	Smc5	NKLES**D**YMAASSQLR	787
Smc5	IQ**D**EK	412	Smc6	**K**NISEIR	719
Smc5	**E**RETLEK	427	Smc6	IEVE**K**SASILDK	862

Pentamer	**Intra-subunit crosslinks (59)**				
Protein 1	***Peptide 1***	***Residue 1***	***Protein 2***	***Peptide 2***	***Residue 2***
Smc6	**K**EENFSSPK	5	Smc6	NA**K**RPR	16
Smc6	NA**K**RPR	16	Smc6	T**K**EQIHQGEER	234
Smc5	DPSSEVPS**K**R	31	Smc5	**K**TSTPSPQPSK	7
Smc5	DPSSEVPS**K**R	31	Smc5	TSTPSPQPS**K**R	17
Smc5	DPSSEVPS**K**R	31	Smc5	**K**STTQK	154
Nse1	**K**EAEQVLQK	144	Nse1	**K**MRK	140
Smc5	EKQLETSC**K**EK	230	Smc5	**K**VELDQYR	911
Smc6	EDYSYIMET**K**ER	230	Smc6	T**K**EQIHQGEER	234
Smc6	LTEL**K**R	248	Smc6	QCVE**K**EER	254
Smc5	QEYEEV**K**LVR	286	Smc5	V**K**EEVRK	293
Smc5	DRV**K**EEVR	293	Smc5	**K**LKEGQIPVTCR	298
Smc5	DRV**K**EEVRK	293	Smc5	L**K**EGQIPVTCR	300
Smc6	**K**MEEQQVR	310	Smc6	RHYEE**K**QK	821
Smc5	ATDIKEASQ**K**CK	337	Smc5	Q**K**QDVIER	341
Smc5	**K**DKHIEELQQALIVK	348	Smc5	Q**K**QDVIER	341
Smc5	KD**K**HIEELQQALIVK	350	Smc5	Q**K**QDVIER	341
Smc6	ADVVA**K**K	356	Smc6	IEAEN**K**YDAIK	779
Smc6	**K**ISWLK	409	Smc6	**K**NISEIR	719
Smc5	ALCEGEIID**K**R	424	Smc5	**K**LQAVDSGLIALR	665
Smc5	ETLE**K**EK	433	Smc5	**K**SVDDHIVR	436
Smc6	QL**K**ELK	467	Smc6	REELDV**K**HALSYNQR	456
Smc5	DKF**K**QR	478	Smc5	VNAVIAP**K**SSYADKAPSR	542
Smc5	L**K**QIYTAEEK	610	Smc5	SSYAD**K**APSR	548
Smc6	DI**K**HNEELLK	697	Smc6	D**K**EEHGK	442
Smc6	CQLHY**K**ELK	711	Smc6	**K**NISEIR	719
Smc5	A**K**LVTELTNLIK	748	Smc5	I**K**EINVQK	740
Smc5	A**K**LVTELTNLIK	748	Smc5	Q**K**QDVIER	341
Smc6	HYEE**K**QKEHLDTLNK	821	Smc6	**K**MEEQQVR	310
Smc6	HYEEKQ**K**EHLDTLNK	823	Smc6	**K**MEEQQVR	310
Smc6	Q**K**IQAEHASHGDR	877	Smc6	SASILD**K**EINR	869
Smc5	EEEIEQLTEEL**K**GK	908	Smc5	**K**VELDQYR	911
Smc5	**D**PSSEVPSK	23	Smc5	TSTPSPQPS**K**R	17
Smc5	**D**PSSEVPSK	23	Smc5	TSTPSPQPS**K**R	17
Smc5	DPSS**E**VPSK	27	Smc5	TSTPSPQPS**K**R	17
Smc5	DPSSEVPS**K**R	31	Smc5	**E**IDVAK	139
Nse3	DGFA**E**EAPSTSR	45	Nse3	SQ**K**QLELK	83
Smc5	**E**IDVAK	139	Smc5	**K**TSTPSPQPSKR	7
Smc5	**E**IDVAK	139	Smc5	KTSTPSPQPS**K**R	17
Smc5	**E**IDVAK	139	Smc5	**K**TSTPSPQPSKR	7
Smc5	**E**IDVAK	139	Smc5	TSTPSPQPS**K**R	17
Nse1	**E**GEFTLHGR	163	Nse1	ESGVL**K**SNK	255
Nse1	EG**E**FTLHGR	165	Nse1	ESGVL**K**SNKK	255
Smc6	LT**E**LK	246	Smc6	SASILD**K**EINR	869

Smc5	I**E**E**M**ENER	312	Smc5	L**K**EGQIPVTCR	300
Smc6	LNEA**E**QK	322	Smc6	Y**K**DIQDKLEK	326
Smc5	ATDI**K**EASQK	332	Smc5	I**E**E**M**ENER	312
Smc6	APECMAL**K**ADVVAK	350	Smc6	IS**E**ETNAR	337
Smc6	I**E**ELK	389	Smc6	**K**STDQSLEPER	393
Smc6	I**E**ELK	390	Smc6	**K**STDQSLEPER	393
Smc6	I**E**ELKK	390	Smc6	KDDEQLC**K**R	386
Smc6	RIEEL**K**K	392	Smc6	ST**D**QSLEPER	396
Smc6	D**K**EEHGKIKR	442	Smc6	**E**ELDVK	451
Smc6	DKEEHG**K**IKR	447	Smc6	**E**ELDVK	451
Smc5	**K**ASTK	734	Smc5	LMEQDTCNLE**E**EER	730
Smc5	**K**ASTK	734	Smc5	LMEQDTCNL**E**EEER	729
Smc5	IKEINVQ**K**AK	746	Smc5	LM**E**QDTCNLEEEER	722
Smc6	EKELEE**K**MSQAR	846	Smc6	**E**TYLDLDSK	900
Smc6	IEV**E**K	861	Smc6	SASILD**K**EINR	869
Smc5	G**K**KVELDQYR	910	Smc5	IQD**E**K	413

* The table reports specific amino-acid positions that were crosslinked during the XL-MS analysis of the Smc5/6 complex. Protein 1 and 2 columns refer to the specific protein(s) that were connected by a specified crosslink. Amino-acids labeled in bold in peptide 1 and 2 columns represent the crosslinked moieties; K-K or K-M in the case of the DSS crosslinker and K-D or K-E in the case of the DMTMM. Residue 1 and 2 columns report the specific amino-acid residues that were crosslinked in each protein.

Table 5.S2. Yeast strains used in this study, Related to STAR Methods.

Figure	Name	Relevant genotype[#]
Figure 5.1	D4672	*MAT**a** pep4::HIS3 bar1::hisG lys2::P_{GAL1}-GAL4::lys2* [*2u TRP1 P_{GAL10}-hSMC5 P_{GAL1}-hSMC6*]
	D5819	*MAT**a** tdh3::kanMX6 lys2::P_{GAL1}-GAL4::lys2* [2μ *LEU2d TRP1 P_{GAL10}-hNSE1 P_{GAL1}-hNSE3*] [2μ *LEU2d URA3 P_{GAL10}-hNSE4a-L-hSMC5 P_{GAL1}-hSMC6*]
Figure 5.2	D6287	*MAT**a**/MAT**α** NSE4/nse4::kanMX6 SMC6/smc6::HIS3MX6*
	D6305	*MAT**a**/MAT**α** NSE4/nse4::kanMX6 SMC6/smc6::HIS3MX6 ura3-1/ura3-1::SMC6-L-NSE4::URA3*
	D6274	*MAT**a**/MAT**α** NSE4/nse4::kanMX6 SMC5/smc5::HIS3MX6*
	D6303	*MAT**a**/MAT**α** NSE4/nse4::kanMX6 SMC5/smc5::HIS3MX6 ura3-1/ura3-1::NSE4-L-SMC5::URA3*
	D4107	*MAT**a***
	D224	*MAT**a** smc5-6::HIS3*
	D6313	*MAT**α** nse4::kanMX6 smc5::HIS3MX6 ura3-1::NSE4-L-SMC5::URA3*
	D6314	*MAT**a** nse4::kanMX6 smc5::HIS3MX6 ura3-1::NSE4-L-SMC5::URA3*
Figure 5.4	D4107	*MAT**a***
	D6425	*MAT**a** smc6-9:NAT*
	D6946	*MAT**a** smc6-R135E:kanMX6*
	D6945	*MAT**a** smc6-R144E:kanMX6*
	D6837	*MAT**a** smc5-6:HIS3MX6*
	D6771	*MAT**a** nse4-4:URA3*
	D6955	*MAT**a** nse4-D261A:kanMX6*
	D7025	*MAT**a** nse3-5:URA3*
	D7016	*MAT**a** nse3-E265R:kanMX6*
	D7024	*MAT**a** nse3-L268K:kanMX6*
	D6796	*MAT**a** nse1-5:URA3*
	D7225	*MAT**a** SMC6-3HA::kanMX6*
	D7224	*MAT**a** smc6-R135E-3HA::kanMX6*
	D7223	*MAT**a** smc6-R144E-3HA::kanMX6*
Figure 5.S1	D4619	*MAT**a** lys2::P_{GAL1}-GAL4::lys2* [2μ *HIS3 P_{GAL10}-hNSE4a*]
Figure 5.S2	D5819	*MAT**a** tdh3::kanMX6 lys2::P_{GAL1}-GAL4::lys2 [2μ LEU2d TRP1 P_{GAL10}-hNSE1 P_{GAL1}-hNSE3] [2μ LEU2d URA3 P_{GAL10}-hNSE4a-L-hSMC5 P_{GAL1}-hSMC6*]
Figure 5.S3	D5819	*MAT**a** tdh3::kanMX6 lys2::P_{GAL1}-GAL4::lys2 [2μ LEU2d TRP1 P_{GAL10}-hNSE1 P_{GAL1}-hNSE3] [2μ LEU2d URA3 P_{GAL10}-hNSE4a-L-hSMC5 P_{GAL1}-hSMC6*]

[#]All yeast strains used in this study were derived from K699/K700 parents or congenic strains.

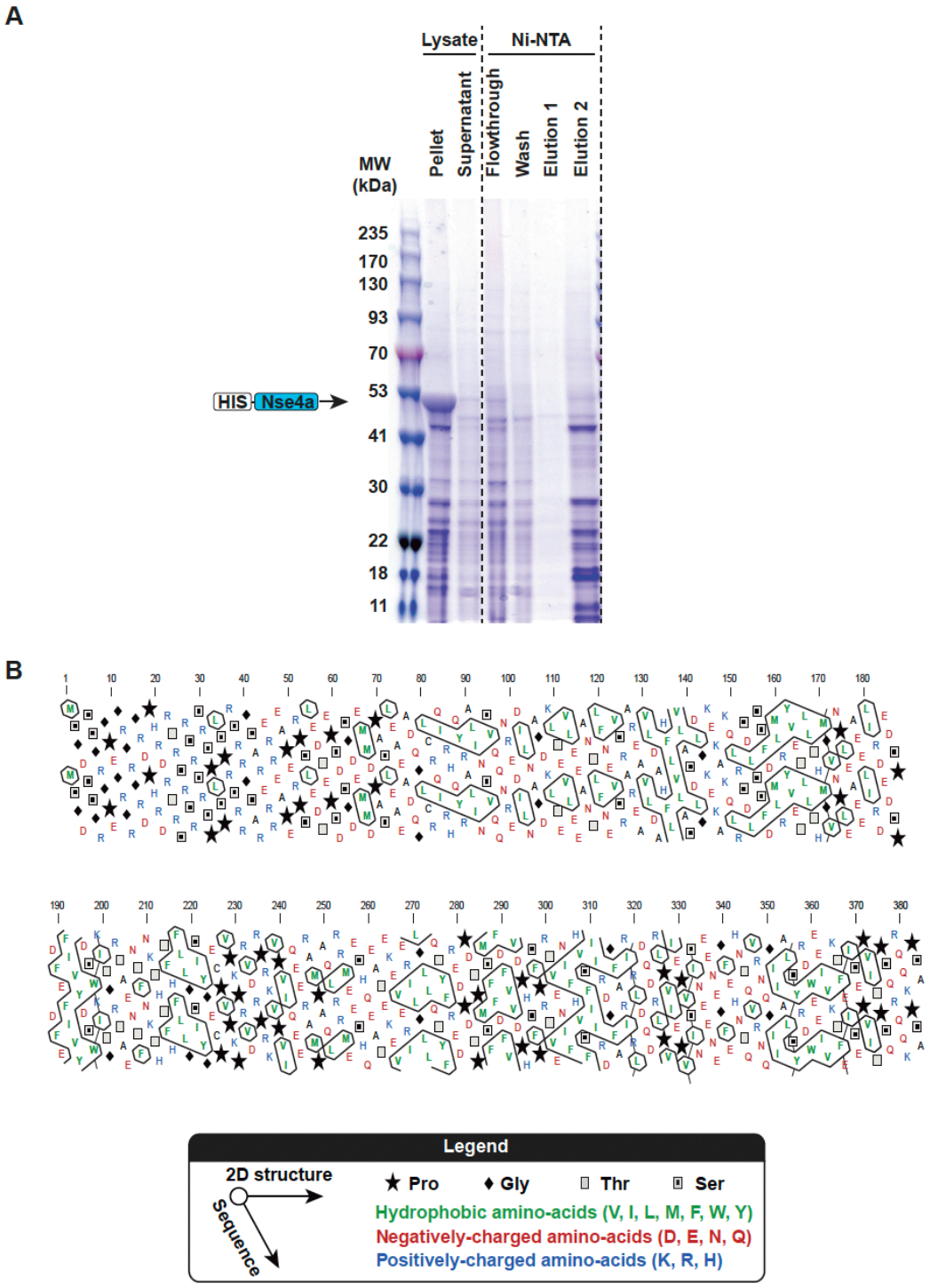


Figure 5.S1. Solubility of His-tagged Nse4a, Related to Figure 5.1. (A) Acrylamide gel showing the various steps involved in the attempt to purify Nse4a. His-tagged hNse4a was overexpressed in yeast and the cell lysate was processed for Ni-chelate chromatography, as previously described (Roy et al., 2011). Nse4a protein could not be isolated/enriched using standard purification procedures.

Importantly, a protein corresponding to the molecular mass of 9xHis-Nse4a became clearly apparent in the cell lysate after overexpression but was found exclusively in the non-soluble fraction of the yeast lysate (*i.e.*, pellet). Various attempts at solubilizing Nse4a using mild detergents or denaturation/renaturation procedures all failed. **(B)** Hydrophobic cluster analysis (HCA) of Nse4a (Bitard-Feildel et al., 2018). HCA version 1.0.2 (from the Mobyle web environment; Ressource Parisienne en BioInformatique Structurale) was used to detect hydrophobic clusters present in Nse4a.

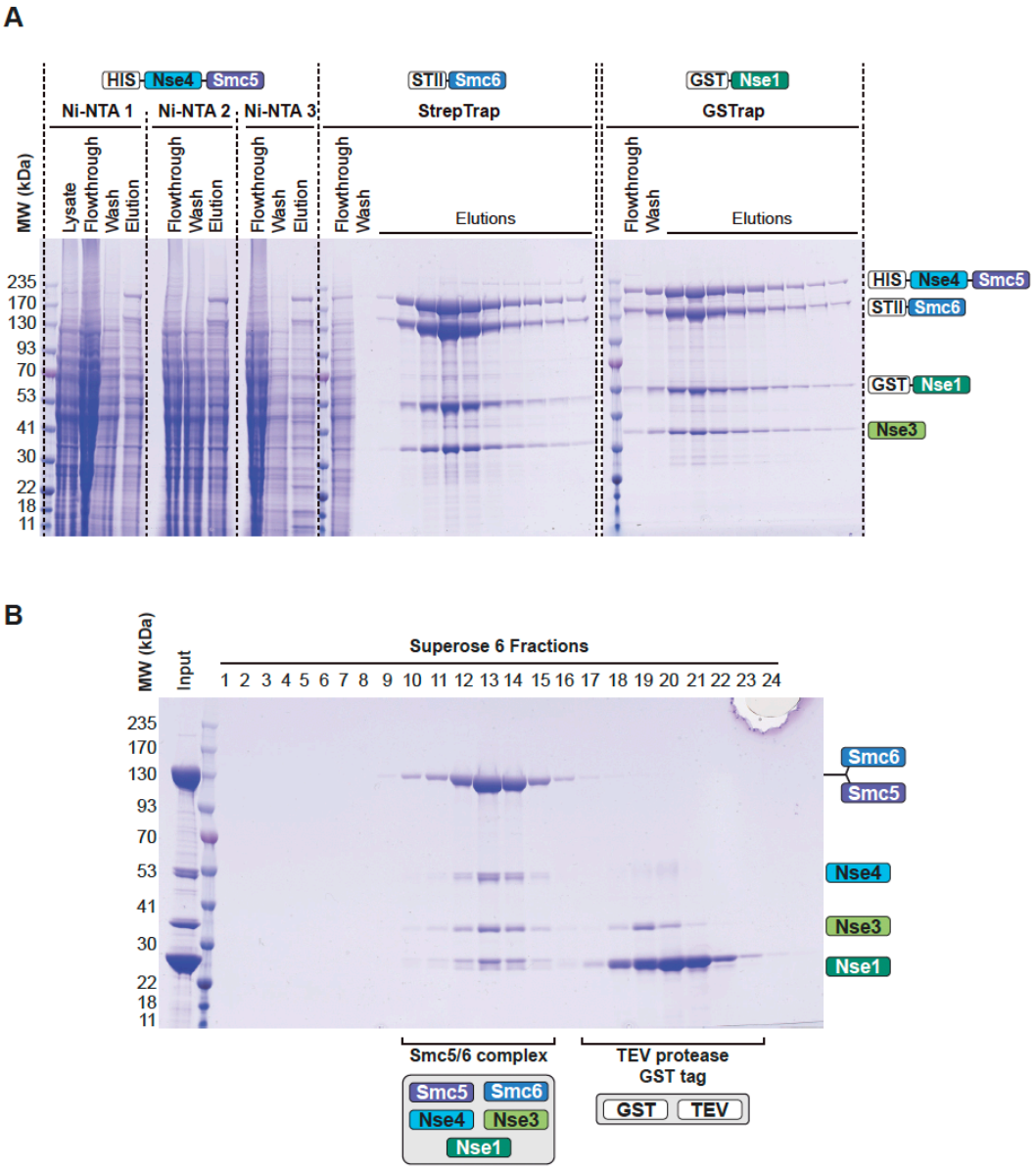


Figure 5.S2. Details of the purification procedure used to isolate the Smc5/6 complex, Related to Figure 5.1. (A) Representative gels showing the purity of the Smc5/6 complex at different steps of Ni-chelate (Ni-NTA), Streptactin Sepharose (StrepTrap) and Glutathione Sepharose (GSTrap) chromatography. The positions of tagged subunits and hNse3 are shown on the right of the gels. Note that the band appearing at ~180kDa prior to TEV cleavage corresponds to the Nse4a-Linker-Smc5 protein. After TEV digestion, Smc5 was released from the fusion and co-migrated with Smc6 on gel (*i.e.*, calculated molecular weights of 128 and 126kDa, respectively). Furthermore, 2 new bands corresponding to Nse4a appeared at ~44kDa after TEV digestion. These bands represent variants of Nse4a that were cleaved at different TEV sites in the linker sequence of the fusion protein (*i.e.*, multiple TEV sites are present in the linker to facilitate cleavage). TEV digestion of GST-Nse1 also gave rise to a novel band in the 25-30kDa range by SDS-PAGE, consistent with the calculated molecular

weight of Nse1 (31kDa), GST (27 kDa), and TEV protease (27 kDa). As shown below, the GST tag and TEV protease were removed from the Smc5/6 core complex using size exclusion chromatography. Specific buffer conditions used for protein binding, wash and elution are described in the Method Details. **(B)** Size exclusion chromatography after TEV-induced removal of purification tags/linker on the subunits of the Smc5/6 complex. The gel shows the co-elution of subunits as a single monomeric Smc5/6 complex during the gel filtration (Superose 6) chromatography step. The positions of untagged proteins are shown on the left side of the gel. Note that the TEV protease and GST tag (from the GST-Nse1 fusion) elute separately and in later fractions than Smc5/6 complex subunits during the Superose 6 chromatography. Detailed conditions used for size exclusion chromatography are provided in Method Details.

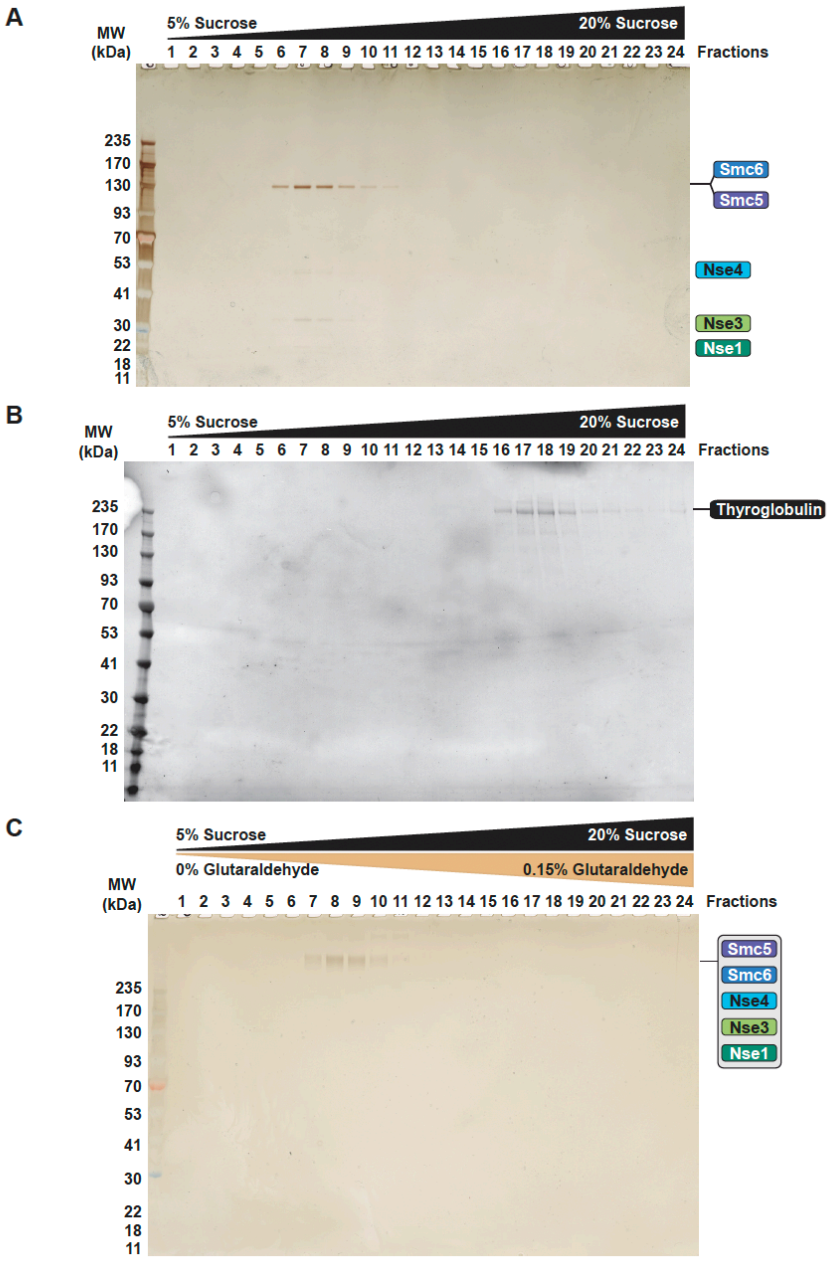


Figure 5.S3. Density gradient centrifugation of the Smc5/6 core complex, Related to Figures 5.1 and 5.3. (A) The sedimentation behavior of the Smc5/6 core complex was monitored by centrifugation in 5–20% sucrose gradients under native conditions. The observed and predicted positions of Smc5/6 complex subunits are shown on the right of the gel. Note that some NSE subunits are weakly stained by silver staining. **(B)** Sedimentation behavior of native thyroglobulin. The distribution of this protein within the sucrose gradient is that of a typical 670kDa monomer. Note how thyroglobulin sediments at a substantially heavier position (fractions 16-20) than the Smc5/6 complex (fractions 6-10). **(C)** Sedimentation behavior of the Smc5/6 core complex under GraFix conditions (*i.e.*, dual 0-0.15% glutaraldehyde / 5-20% sucrose gradients).

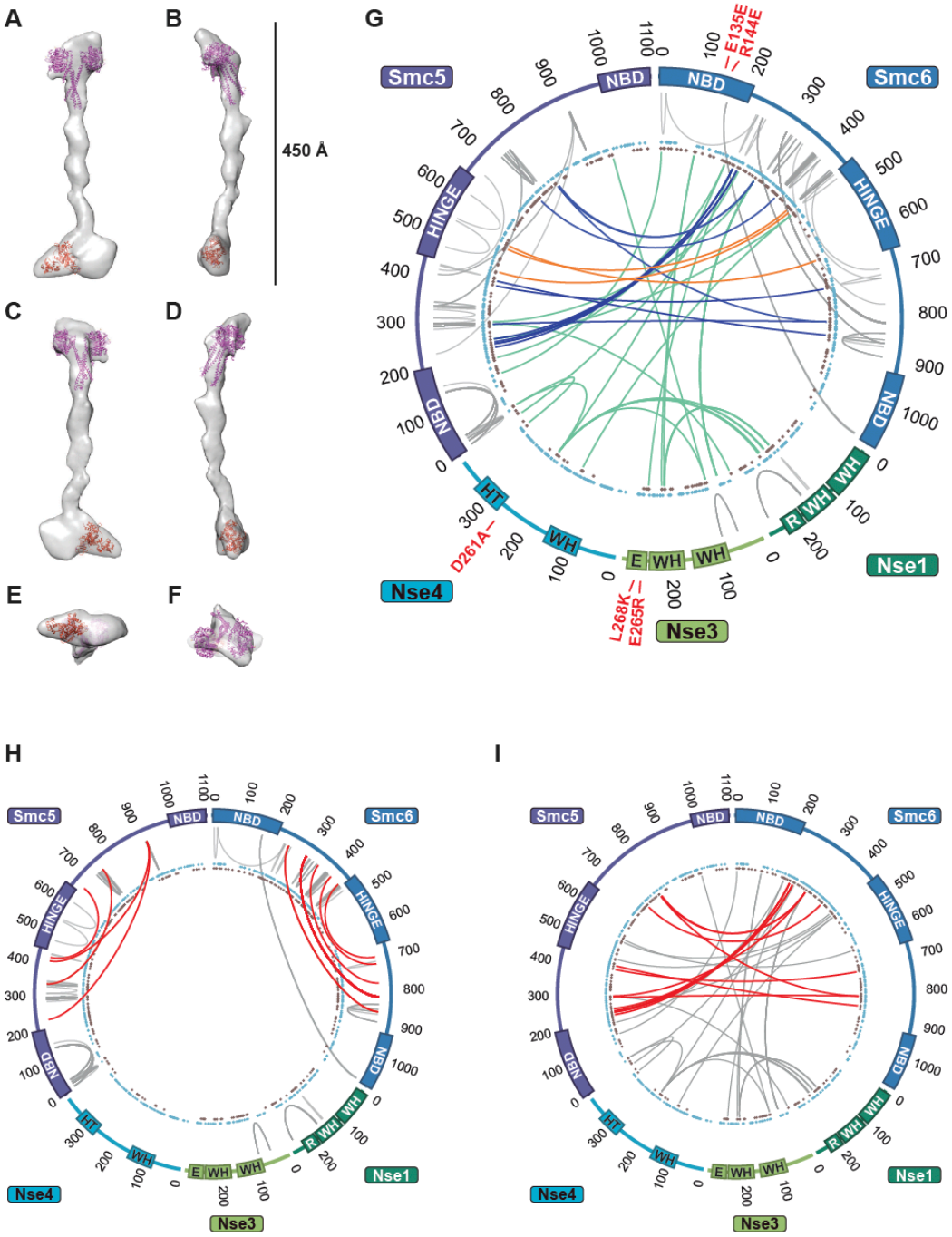


Figure 5.S4. Structural features of the Smc5/6 complex deduced from modeling and subunit proximity networks, Related to Figure 5.3. (A-F) Three-dimensional modeling of the Smc5/6 core complex. Two-dimensional average maps of the Smc5/6 complex were used to generate an initial three-dimensional model of the complex with Xmipp's RANSAC algorithm (Vargas et al., 2014) within the Scipion program suite (de la Rosa-Trevin et al., 2016). The model was then refined in RELION 3.0 (Scheres, 2012). For visualization purposes, spurious noise was removed with the "Hide Dust" feature in UCSF Chimera (Pettersen et al., 2004). Crystal structures of Nse1/Nse3 (PDB: 5WY5; Doyle et al.,

2010) and fission yeast Smc5/6 hinge domain (PDB: 5MG8; Alt et al., 2017) were docked into the three-dimensional map at their known positions within the Smc5/6 complex. This map provides a low-resolution (~52 Å) overview of the scale and general 3D organization of the Smc5/6 complex. Overall shape and appearance of the Smc5/6 core complex at perpendicular angles **(A-D)**, as well as bottom **(E)** and top **(F)** views. **(G-I)** Proximity maps showing interactions identified by XL-MS within the Smc5/6 core complex. **(G)** Network of inter- and intra-subunit crosslinks identified in the Smc5/6 core complex. This schematic highlights the relative position of mutations analysed in this study (marked in red) on a magnified version of the diagram shown in Figure 5.2H. **(H)** Interaction map highlighting intra-molecular connections involving the SMC arms of the Smc5/6 complex (red lines). **(I)** Network of inter-molecular crosslinks connecting the arms of different SMC subunits in the Smc5/6 complex (red lines). Other linkages are labelled in grey.

A

	Smc5/6 Dimer		Smc5/6 Complex	
	Kd values (nM) [R square]	Hill slope	Kd values (nM) [R square]	Hill slope
ssDNA	32.42 ± 1.909 [0.9821]	2.475 ± 0.2554	27.32 ± 0.7638 [0.9901]	3.310 ± 0.2624
dsDNA	121 ± 7.482 [0.9639]	4.427 ± 0.7562	81.81 ± 3,048 [0.9243]	4,787 ± 0.7145

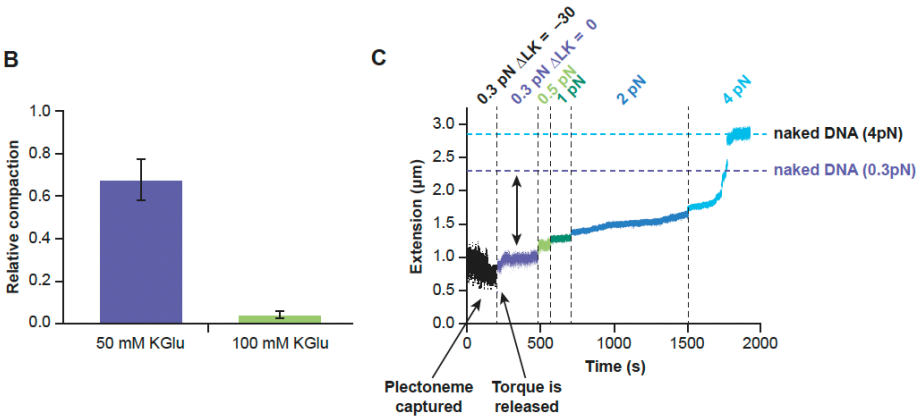


Figure 5.S5. Behavior of the Smc5/6 complex when interacting with DNA substrates under various experimental conditions, Related to Figures 5.5-5.7. (A) Smc5/6 dimer and complex affinity constants and Hill coefficients observed for ss and dsDNA binding reactions shown in Figure 5.5. **(B)** DNA compaction at ΔLk = 0 depends on potassium glutamate (KGlu) concentration. For a series of 5 experiments with 5 nM Smc5/6 core complex and with DNA held at 0.3 pN with no torsional stress (ΔLk = 0), the amount of compaction (fractional reduction in extension of tether) was nearly 70% in 50 mM KGlu and 1 mM ATP, and less than 5% (essentially zero) in 100 mM KGlu and 1 mM ATP. Error bars indicate SEM for N=5 measurements. **(C)** Supercoiled DNA "captured" by the Smc5/6 core complex is stable against applied forces up to 2 pN. In a plectonemically-supercoiled DNA "capture" experiment similar to that shown in the main text, supercoiled DNA was captured and then linking number was returned to ΔLk = 0 (black points before time of 200 sec) and then force was gradually increased (first increase to 0.5 pN at time of 500 sec). The horizontal purple line shows the initial (naked, pre- Smc5/6 core complex) DNA length at 0.3 pN. The DNA extension lags behind this level for ~1000 sec until a large force of 4 pN is applied, which in 500 sec returns DNA extension to its naked level for 4 pN (horizontal black line) via a series of extension jumps.