THE ROLE OF SON IN CHROMATIN-MEDIATED GENE EXPRESSION

A Dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

by

MELISSA JORDAN WARD

B.S., Wright State University, 2016

2022

Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

April 28, 2022

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Melissa Jordan Ward ENTITLED The role of SON in chromatin-mediated gene expression BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

_____

Paula Bubulya, Ph.D.
Dissertation Director

_____

David Ladle, Ph.D.
Program Director, Biomedical Sciences Ph.D. Program

_____

Barry Milligan, Ph.D.
Vice Provost for Academic Affairs
Dean of the Graduate School

Committee on Final Examination:

_____

Michael Leffak, Ph.D.

_____

Weiwen Long, Ph.D.

_____

Labib Rouhana, Ph.D.

_____

Quan Zhong, Ph.D.

ABSTRACT

Ward, Melissa Jordan. Ph.D., Biomedical Sciences Ph.D. Program, Wright State University, 2022. The role of SON in chromatin-mediated gene expression.

Gene expression in mammalian cells requires complex nuclear choreography, and there is increasing evidence that spatiotemporal organization of chromatin and nuclear compartments plays an important role in gene expression. In this dissertation, I examined the function of SON, a splicing factor with a known role in nuclear organization, in chromatin-mediated gene expression. SON association with a transcriptionally inactive U2OS 2-6-3 reporter gene array provided a useful model to study SON's chromatin dynamics. I demonstrated that SON associates with the inactive but not the activated array, and that SON's RNA binding domains are not necessary for that association. Second, I discovered a new role for SON in maintaining chromatin condensation. Whereas chromatin decondensation is typically correlated with transcription activation, I have demonstrated that reporter transcripts are not produced at decondensed SON-depleted U2OS 2-6-3 reporter gene loci, and that SON-depleted loci contain histone H3 that is trimethylated on lysine 9, a marker for transcriptionally silent chromatin. We found that SON depleted reporter loci are still transcriptionally activatable, and that inhibiting transcription elongation is not sufficient to condense the enlarged SON depleted reporter locus. These findings suggest that higher order chromatin structure and transcription activation are functionally distinct mechanisms of gene regulation that can

be uncoupled. Finally, I investigated SON's role in genome-wide chromatin organization. SON-depleted cells are more susceptible to DNase digestion, implicating SON in the maintenance of chromatin stability globally. In conclusion, this study demonstrates a new function for the splicing factor SON in maintaining chromatin organization.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor, Dr. Paula Bubulya, for her unwavering encouragement and support all these years together. None of this would have been possible without her expertise and enthusiasm. I would also like to thank each member of my dissertation committee: Dr. Michael Leffak, Dr. Weiwen Long, Dr. Labib Rouhana, and Dr. Quan Zhong for their time, guidance, and stimulating questions.

Thank you to all the current and former members of the Bubulya lab, and Dr. Athanasios Bubulya and Jim Frisbie for their camaraderie and guidance. Thank you to the BMS program for their support and this opportunity. I'd like to thank Michael Bottomley at the Statistical Consulting Center for his very patient explanations, and the NIH as the funding agent for this project. A sincere thank you to all the educators in my career that have been such wonderful advocates for higher learning and women in STEM.

Thank you to my friends for all their love and support. Dr. XiuHuan Yap, Jananie Rockwood, Dr. Sara Seibert, Dr. Hannah Shows, Christian Garrido, Jaylen Hudson, Chris Waker, Briana Childers, and Matt Van Voorhis. I would like to extend a very special thank you my family Douglas Ward, Joshua Ward, and Jacob Ward, for their steadfast love and encouragement. In particular, thank you to my mom Tammy Ward; words are not enough to express my love and gratitude. And to my cat Pebbles, my ever-faithful supporter.

DEDICATION

In loving memory of my grandparents, Don and Peggy Taylor, and Richard and Georgiana Ward.

CHAPER I

INTRODUCTION

Mammalian Cell Nucleus Organization

Eukaryotic cells are highly compartmentalized, with spatially separated organelles executing cellular functions in a dynamic, 4-dimensional space. Lipid membranes physically separate many of these organelles to create microenvironments. These microenvironments have distinct compositions that allow for biochemical gradients and increased availability of organelle-specific factors, thereby increasing the efficiency of biochemical reactions. The eukaryotic cell nucleus in particular is an extraordinarily complex and dynamic membranous cellular organelle that contains nearly all of the cell's DNA and the factors necessary for gene expression.

Phase Separation

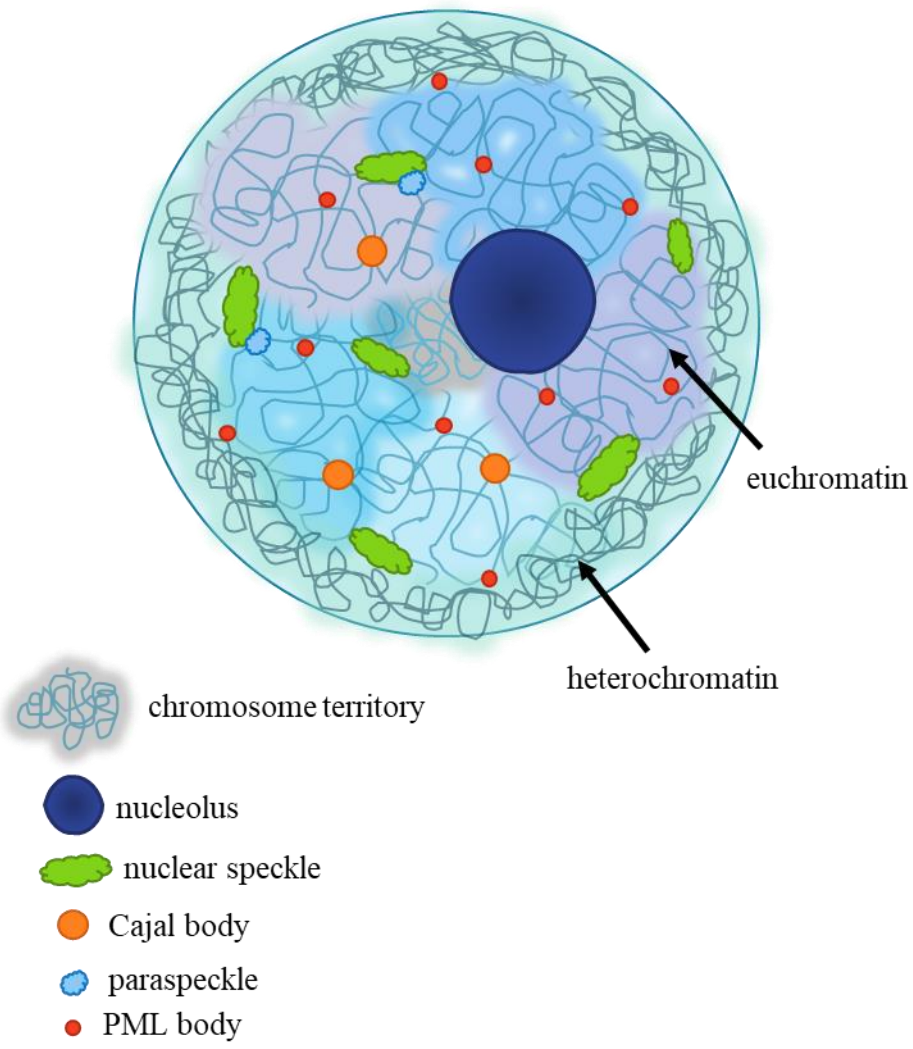
In humans and other higher order eukaryotes, the cell nucleus is further compartmentalized by many non-membranous bodies that self-associate by phase separation and are the storage and assembly sites for factors involved in major processes such as transcription, DNA replication, and cell cycle regulation. The largest nuclear subcompartment, nucleoli, was observed as early as the 1800's (Wagner, 1835). Understanding the process by which these "biomolecular condensates" self-associate and

their involvement in gene expression has gained ever increasing popularity in the cell biology and biophysics realms over the past 20 years (Banani et al., 2017; Boeynaems et al., 2018 and Lesne et al., 2019). The simplest example of phase separation is the nonmixing of oil and vinegar or cornstarch and water, where aggregates called colloids form to visibly separate immiscible substances. In cells, concentrated biomolecules including proteins and RNAs are thought to self-associate to form droplets via low-entropy interactions between low-complexity (more homogenous) amino acid regions (Brangwynne et al., 2009; Feric et al., 2016; Hennig et al., 2015; Hondele et al., 2019).

Much like how oil and vinegar can be mixed by shaking or unmixed by allowing to settle, phase separated nuclear bodies are dynamic and can associate or dissociate based on cell cycle phase, stress response, or modulation in gene expression (Alberti et al., 2020; Nott et al., 2015; Phair and Mistelli, 2000). Phase separation has been proposed as a mechanism in the formation of several types of nuclear nonmembranous subcompartments including nucleoli, paraspeckles, PML bodies, Cajal bodies, and nuclear speckles (Sawyer et al., 2018). These nuclear bodies have increased concentrations of functionally distinct biomolecules relative to the nucleoplasm, effectively forming storage and assembly sites for factors involved in processes such as ribosomal biogenesis (nucleoli), stress response (PML bodies), histone gene synthesis (histone locus body), snRNP assembly (Cajal bodies), and mRNA processing (nuclear speckles and paraspeckles). A cartoon of selected nonmembranous bodies in mammalian cell nuclei is depicted in Figure i.

Figure i. Cartoon depiction of mammalian cell nucleus organization.

The mammalian cell nucleus contains a variety of nonmembranous bodies and higher order organization in space and time. Tightly compacted transcriptionally inactive heterochromatin tends to localize to the nuclear periphery (green), while more active and decondensed euchromatin organize to the nuclear interior. Interphase chromosomes, although less restricted than during mitosis, are spatially separated into chromosome territories. The nucleolus is a large and rounded granular structure functioning in ribogenesis (dark blue). Nuclear speckles reside in gene-poor regions of the nucleus and are the storage and assembly sites for pre-mRNA processing factors (green). Paraspeckles localize near speckles and function in RNA retention (blue). Cajal bodies function in snRNP assembly (orange), and PML bodies are localized near DNA loops and involved in cellular stress response (red).


euchromatin
heterochromatin
chromosome territory
nucleolus
nuclear speckle
Cajal body
paraspeckle
PML body

Nuclear Speckles

Nuclear speckles are the storage and assembly sites for mRNA processing factors. Interphase nuclei have 20-50 nuclear speckles that range from 0.5 μm to a few microns in diameter and exhibit a granular morphology (Spector and Lamond, 2011). They are typically found near active transcription sites, although little to no DNA or mRNA has been detected inside the speckle core (Galganski et al., 2017; Mintz et al., 1999; Saitoh et al., 2004; Spector and Lamond, 2011). Disruption in speckle assembly or function has disastrous effects on pre-mRNA processing and disease due to splicing malfunction (Tripathi et al., 2012).

Speckles contain both protein and non-coding RNA (ncRNA) components and are reliably identified by the presence of SRSF2 (also known as SC35) protein (Fei et al., 2017). Speckles have an inner core of SRSF2 and SON, and an outer coat of MALAT1 lncRNA and small nuclear ribo-nuclear proteins (snRNPs) (Fei et al., 2017; Sun et al., 2018, Tripathi et al., 2010). This proteinaceous inner core has led to some interest in potential liquid-liquid phase separation (LLPS) properties, and recent studies suggest that LLPS may enhance nuclear speckle nucleation (Lesne et al., 2019; Nott et al., 2015; Banani et al., 2017; Uversky, 2016). This phase separation is perhaps due to the abundance of low-complexity regions (LCRs) in nuclear speckle proteins (Galganski et al., 2017; Marzahn et al., 2016), although the exact assembly mechanism of nuclear speckles remains unclear (Mao et al., 2011).

Similar to other nonmembranous organelles, nuclear speckle organization is highly dynamic, with a constant flux of factors despite appearing positionally stable throughout interphase by microscopy (Phair and Mistelli 2000). Storage of pre-mRNA

processing factors in nuclear speckles regulates the availability of splicing factors at transcription sites. Speckle-associated SR proteins are stored until phosphorylation by Clk1 kinase for recruitment to active transcription sites (Bubulya et al., 2004; Misteli et al., 1998; Sacco-Bubulya and Spector, 2002). In fact, when Clk1 is overexpressed, nuclear speckles are abolished, and splicing factors disbursed throughout the nucleus (Sacco-Bubulya and Spector, 2002). Increased availability of splicing factors via nuclear speckle disbursement is associated with faster splicing kinetics and decreased mRNA retention at active transcription sites (Hochberg-Laufer et al., 2019).

Although splicing does not occur to a large extent inside speckles, chromatin regions containing highly transcribed genes are located near the nuclear speckle periphery (Brown et al., 2008; Hu et al., 2010; Quinodoz et al., 2018; Yang et al. 2011). Positional analysis of housekeeping genes by chromatin conformation capture (3C) has shown that housekeeping genes can cluster into large gene-expression hubs that interact with nuclear bodies and drive transcription. Newer evidence has emerged that heat shock proteins are clustered around the edge of nuclear speckles, and that dissociation of speckles results in a decrease in transcription at these gene expression hubs. Chen et al. (2018) found that "transcriptional hot zones," including housekeeping genes, early replicating regions, genes with low transcriptional pausing, and super-enhancers are more likely to be positioned near speckles (Speckle Associated Domains; SPADs) in K562 cells. Furthermore, disruption of nuclear speckle assembly by knockdown of Srrm2 reduced the overall number of chromatin interactions in topologically associated domains (Hu et al., 2019; Ilik et al., 2020). Therefore, the spatiotemporal positioning of speckles

may be closely tied to the surrounding chromatin structures and provide a proximity hub to increase the efficiency of pre-mRNA processing of actively transcribed genes.

3D Chromatin Organization

The interphase eukaryotic cell nucleus contains about 2 meters of DNA that is highly organized and compacted. 147 bp of DNA are wrapped 1.7 negatively supercoiled turns around the core nucleosome, with a 10-80 bp variable linker DNA between nucleosomes. This DNA-nucleosome complex forms the 10 nm fiber, also considered the "beads on a string" model where the beads are the nucleosomes and string is the linker DNA. This fiber can be further condensed into a helical 30 nm fiber, and supercompacted into chromosomes in mitosis (Hubner et al., 2013; Zidovska, 2020).

A model for how spatial positioning of chromatin fibers can affect gene expression is the lampbrush chromosome in amphibian oocytes (Callan, 1986). In lampbrush chromosomes, active chromatin regions are decondensed and loop outwards, while transcriptionally inactive chromatin is highly condensed at internal centromeric regions. Another example is polytene chromosomes in Drosophila. In this model, the 30 nm chromatin fiber is organized by looping along a protein scaffold. When genes are activated, these chromatin loops puff outward, causing gene specific decondensation, which increases chromatin accessibility and promotes gene expression. Polytene chromosomes are large enough that chromatin organization can be visualized by light microscopy. Darker banding of the chromosomes corresponds to regions of highly condensed, silent chromatin, while the lighter corresponds with the transcriptionally active gene puffs. The transitions between condensed and decondensed chromatin is

dynamic, and the transcriptionally active puffs eventually recede. The lampbrush and polytene chromosomes describe a chromatin structure where chromatin decondensation is coupled with transcription activation (Callen, 1986; Morgan, 2002).

There has been a massive effort to improve our understanding of higher order chromatin structure in the last few decades. The development of the 3C technique has been especially critical in mapping the chromatin organization (Dekker et al., 2002). In 3C, crosslinked chromatin is subjected to restriction enzyme digestion and proximity ligation to generate chimeric biomolecules that are sequenced and function as evidence of intra-chromatin interactions (McCord et al., 2020; Tena and Santos-Periera, 2021). This method can be applied genome-wide in high-resolution (Hi-C), generating a map of short-range and long-range chromatin interactions (Belton et al., 2012).

Understanding which chromatin regions interact in 4D space and the effect of chromatin organization in gene expression and disease is the subject of the NIH-funded 4D Nucleome [4DN] consortium. 4DN was created in 2014 and is a global network of researchers working to further our understanding of 3D nuclear organization in space and time (4D Nucleome, 2021). To date, the consortium has published over 600 papers and developed dozens of new software packages and methodologies. As part of this effort, some of the classical models of nuclear organization have been updated to include new descriptive terms of 3D structure, which I will be describing here.

In interphase nuclei, chromatin fibers are spatially organized into loops called topologically associated domains (TADs). Chromatin fibers are not arranged in a linear sequence, but rather organize in clusters with other loops that may be kilobases away in 2D space but associate together in 3D nuclear space. TAD boundaries are enriched in

housekeeping genes, tRNA and the protein CTCF (Anania and Lupianez, 2020; Wang et al., 2021). Transcriptionally active and inactive TADs are designated as compartment A or compartment B chromatin, respectively, and are further organized in radial space (Chung et al., 2016). Compartment B chromatin is generally organized at the nuclear periphery (lamina associated domain: LAD), while Compartment A chromatin is positioned near the nuclear interior (nucleolus associated domain: NAD). At the highest level of interphase organization, whole chromosomes occupy distinct nuclear territories (Cremer and Cremer, 2010).

The phrase "structure determines function" is especially relevant when thinking of higher order chromatin organization. A recent report found that TAD boundaries are more evolutionarily conserved than TADs themselves, and TAD boundaries tend to be cell-type specific (McArthur and Capra, 2021). Depletion of boundary proteins CTCF or cohesin decreased the intensity of intra-TAD interactions without affecting TAD nuclear positioning (Zuin et al., 2013). These reports support a chromatin organization model where short-range TAD interactions are dynamic and flexible, while long-range interactions between TADs are more tightly regulated with more stable boundaries. In acute myeloid leukemia (AML), an inversion on chromosome 3 leads to TAD breakdown, and the formation of two new TADs. This nuclear repositioning results in both haploinsufficiency of GATA-2 which is dysregulated in AML, and misplacement of the GATA-2 enhancer near the EVI oncogene, contributing to tumor formation (Dekker and Heard, 2015; Groschel et al., 2014). These studies show that chromatin fibers are nonrandomly positioned, cluster into higher-order functional domains, and suggest that perturbation in 3D chromatin organization can contribute to disease.

SON protein

SON structure

SON is a large (250 kD) protein that was first identified in proteomic analysis of nuclear speckles, formerly known as interchromatin granule clusters (Saitoh et al., 2004; Takada et al., 2009). SON is evolutionarily conserved in higher eukaryotes. A smaller SON sequence containing the C-terminal G-patch and RNA recognition motifs is highly conserved in non-mammal vertebrates, whereas the N-terminus and repeat domain is specific to mammals (Wynn et al., 2001). SON is ubiquitously expressed in human tissues, although higher SON expression has been reported in hematopoietic organs (Ahn et al., 2013) and undifferentiated embryonic stem cells (Hoffman et al., 2008). The conserved stable expression of SON in human tissues supports a role for SON in cellular maintenance and stability (Hickey et al., 2014).

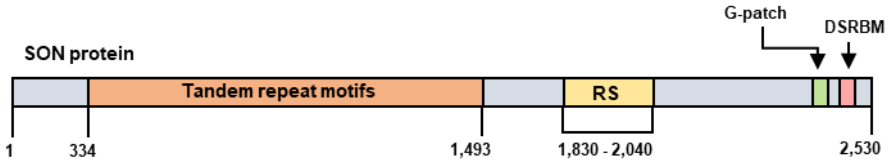
Our laboratory was the first to identify SON's role in nuclear speckle organization. Sharma et al., (2010) examined SON's function in nuclear speckle integrity and cell-cycle regulation, using fluorescent microscopy to identify endogenous SON localization and examine the effect of SON depletion on cell integrity. Instead of round, regular speckles, the SON-depleted speckles appear distinctly donut-shaped, resulting in a reorganization of several nuclear speckle proteins including U1-70K and SRSF1. Subsequent studies have further identified that SON is localized to the core of nuclear speckles with SC35 (Fei et al., 2017; Hu et al., 2019; Ilik et al., 2020), and that SON is required for serine/threonine kinase CK1 localization to speckles (Kuga et al., 2016), suggesting a role for SON in speckle seeding and/or scaffolding.

A schematic diagram of SON sequence motifs is shown in Figure ii. About a third of SON's amino acid (AA) sequence consists of a unique tandem repeat domain that is still relatively uncharacterized. The repeat domain is required for nuclear speckle organization (Sharma et al., 2010) and recruitment of SON to the reporter BTM locus (Battini et al., 2010). More studies are needed to biochemically characterize these repeats and the role they may play in nuclear organization. Downstream of the repeat motifs, SON contains a series of RNA recognition motifs, including a G-patch, double-stranded RNA binding motif (DSRBM) and an RS domain that is a hallmark of a class of mRNA processing factors called SR proteins. Our laboratory previously developed a C-terminal SON antibody (WU13) to examine endogenous SON dynamics, which recognizes the full-length SON isoform (Sharma 2011).

Truncated SON isoforms have also been identified, including SON-B and SON-E (Kim et al., 2016). SON-B is missing the C-terminal RNA recognition domains DSRBM and G-patch, while SON-E is missing these RNA-binding domains as well as the SR domain. In normal conditions, full-length SON-F represses target gene promoters and prevents mixed lineage leukemia (MLL) complex assembly. However, in leukemic conditions, the shorter SON-B and SON-E isoforms are over expressed and outcompete SON-F to associate with target gene promoters. The shorter isoforms are not capable of repressing MLL complex, resulting in de-repression of multiple leukemia-associated genes (Kim et al., 2016).

Figure ii: Schematic of SON protein domains

The SON amino acid sequence contains multiple functional domains. Tandem repeat motifs span from residues 334 – 1493. An RS domain spans residues 1830 – 2040. At the C-terminal end is a G-patch and double stranded RNA binding motif.



SON in splicing

SON is the largest SR protein, and as such, contains a serine/arginine-rich domain (RS domain). SR proteins are a class of proteins that have an RS domain, classified as a domain that is arginine and serine-rich. Phosphorylation of serines in the RS domain releases SR proteins from nuclear speckles and promotes assembly into splicing complexes on nascent pre-mRNAs (Tripathi et al., 2012). SR proteins regulate alternative splicing in a concentration and phosphorylation-dependent manner (Tripathi et al., 2010). The RS domain suggests regulation of SON functions by phosphorylation. SON also contains a G-patch and RNA binding motifs that function in weak splice-site detection and a putative DNA-binding domain, although the exact sequence is unknown.

SON has been identified as a constitutive protein in the regulation of weak splice sites, specifically requiring the C-terminal RS domain and G-patch. In Ahn et al. (2011), reporter minigenes mimicking weak splice sites were downregulated following SON depletion. Moreover, when these minigenes were altered to mimic strong splice sites, the exon-intron junctions were not affected by SON depletion, suggesting SON is critical in the regulation of weak splice sites. SON also facilitates the recruitment of SR protein SC35 to RNAPII complexes (Ahn et al., 2011). SON knockdown results in the downregulation of genes primarily involved in DNA replication/repair and cell cycle regulation, suggesting SON is essential for the regulation of these processes (Sharma et al., 2011). It is currently unknown if SON's role in splicing regulation is separate from its chromatin stability and nuclear speckle organization functions.

Spatial coordination of SON-mediated gene expression

SON facilitates splicing through SR protein recruitment to splice sites, and SON depletion results in increased splicing defects. However, the exact mechanism of SON-dependent splicing is unknown. A relatively new class of proteins named splicing adapters has been shown to regulate pre-mRNA processing through chromatin association and modification (De Almeida and Carmo-Fonseca, 2012; Matveeva et al., 2016). Namely, these adaptors can read chromatin marks and interact with splicing machinery (Agirre et al., 2015; Luco et al., 2010; Luco et al., 2011). Given SON's role in splicing factor recruitment and regulatory chromatin association, SON may alter the chromatin landscape of target genes as a mechanism for splicing regulation. Indeed, SON binding sites near transcription start sites (TSS) are closely associated with the activating histone modification H3K4me3, with high peaks of H3K4me3 corresponding with the valleys of SON peaks, and vice versa (Kim et al., 2016). Preliminary studies from our laboratory show SON colocalizes with a transcriptionally inactive gene locus and is removed upon transcription activation (Sharma 2011). A pilot SON-ChIP-seq study in our laboratory shows a preference for SON binding to promoters (unpublished), corroborating the findings in Kim et al. (2016). Understanding how SON's putative chromatin association effects downstream gene regulation is a long-term goal of our laboratory.

SON in disease

Given the importance of SON in essential cellular processes like cell cycle regulation, pre-mRNA processing, and nuclear organization, it is no surprise then that SON is dysregulated in many diseases. SON is overexpressed in glioblastoma multiforme (GBM), and knockdown of SON causes intron retention of the oncogenic splicing regulator PTBP1 and inhibits the proliferation and clonogenicity of GBM cells in vitro (Kim et al., 2021). In pancreatic cancer, SON knockdown suppressed pancreatic cancer cell proliferation and was associated with increased survival (Furakawa et al., 2012). Most notably, the newly identified neurodevelopmental disorder Zhu-Tokita-Takenouchi-Kim (ZTTK) syndrome is characterized by de-novo heterozygous loss-of-function (LoF) mutations in SON (Kim et al., 2016b). Affected individuals present with facial dysmorphisms, short stature, and intellectual disability, along with significant brain malformations and musculoskeletal abnormalities. SON LoF occurred via de novo frameshift mutations in exon 3 that were mostly localized to the RS domain and the repeat domain motifs closest to the RS domain (Kim et al., 2016b). Consistent with SON's RS domain being essential for splicing regulation, splicing defects were observed in individuals with SON LoF mutations, including exon skipping in HDAC6 and ADA mRNA as previously reported by our laboratory (Battini et al., 2013; Sharma et al., 2011). The ZTTK SON-Shine Foundation has been created as a hub for ongoing ZTTK research efforts and community support (ZTTK-SON-Shine Foundation, n.d.).

SON has also been implicated in another neurological disorder, Alzheimer's. The well-known tau aggregates that are a hallmark of Alzheimer's disease have shown LLPS properties in the nucleus and/or cytoplasm. Tau aggregates have been reported to localize

to nuclear speckles and carry speckle factors including core speckle protein Srrm2 into the cytoplasm, disrupting pre-mRNA processing (Lester et al., 2021). Furthermore, tau accumulation at nuclear speckles alters their spatial positioning and creating speckle subdomains. SON was found to reorganize to the speckle periphery when tau was aggregated at speckles, which is striking since SON typically localized to the stable speckle core. In neuronal cells, hyperphosphorylated tau increases LLPS, increasing the likelihood of tau aggregation (Wegmann et al., 2018). These data place SON at the intersection of nuclear speckle organization, phase separation, and spatial positioning in tau-aggregated cells, and further highlights the importance of nuclear integrity in disease. Similar droplet-to-aggregate transitions have also been identified in the Parkinson's associated protein α-synuclein, and the amyotrophic lateral sclerosis (ALS) associated nuclear protein FUS and TDP-43 (Pakravan et al., 2021 and Zbinden et al., 2020). SON's involvement in a broad range of cancer types and diseases suggests that SON is a master regulator that is essential for cell health and gene expression regulation.

U2OS 2-6-3 reporter gene array

The U2OS 2-6-3 reporter gene array is a useful model to investigate SON's role in chromatin-mediated gene expression. The U2OS 2-6-3 reporter gene array is stably incorporated into chromosome 1p36 of U2OS osteosarcoma cells, and there are approximately 200 copies at the integration site (Janicki et al., 2004). A schematic of the inducible U2OS 2-6-3 reporter gene construct is shown in Figure iii. The 5' end of the construct contains 256 copies of lac operator sequence that serve as a docking site for a

stably incorporated Lac Repressor fusion protein. Unless otherwise indicated, I used U2OS 2-6-3 cells that have a stably incorporated mCherry-LacI, which I will refer to as the "mU2OS" cell line. The Lac Repressor fusion protein binds to the Lac Operator repeats, allowing us to visualize the transgene array. When transcription is off, the chromatin is condensed (i.e., closed), and the mCherry-LacI is detected as a small (0.89 ± 0.12 μm diameter) circular dot by microscopy (Janicki et al. 2004). When transcription is on, the chromatin is decondensed (i.e., open), and the mCherry-LacI is detected as an expanded and irregular-shaped spot (~2.0 - 4.5 μm diameter) by microscopy (Janicki et al. 2004). Thus, we can determine the chromatin conformation status of the U2OS 2-6-3 gene locus by microscopy using the Lac system. Immediately downstream of the lac operator repeats, 96 copies of the tetracycline response element (TRE) allow for inducible transcription of the reporter gene. When transfected with tetracycline-controlled transactivator (rTetR fused to the VP16 activation domain; pTet-On) in the presence of doxycycline (dox), pTet-ON binds to the TREs and induces transcription by activation of a downstream CMV minimal promoter. The transcribed RNA encodes a cyan fluorescent protein with a peroxisomal targeting sequence (CFP-SKL). When translated, CFP-SKL allows for visualization of protein product when transcription is activated. The reporter gene also codes for MS2 repeats, which, when transcribed, form stem-loops in the reporter mRNA that can be visualized by expression of MS2-coat-YFP fusion protein (YFP-MS2BP). The gene also includes a beta-globin intron-exon cassette to ensure proper pre-mRNA processing and a cleavage/polyadenylation signal.

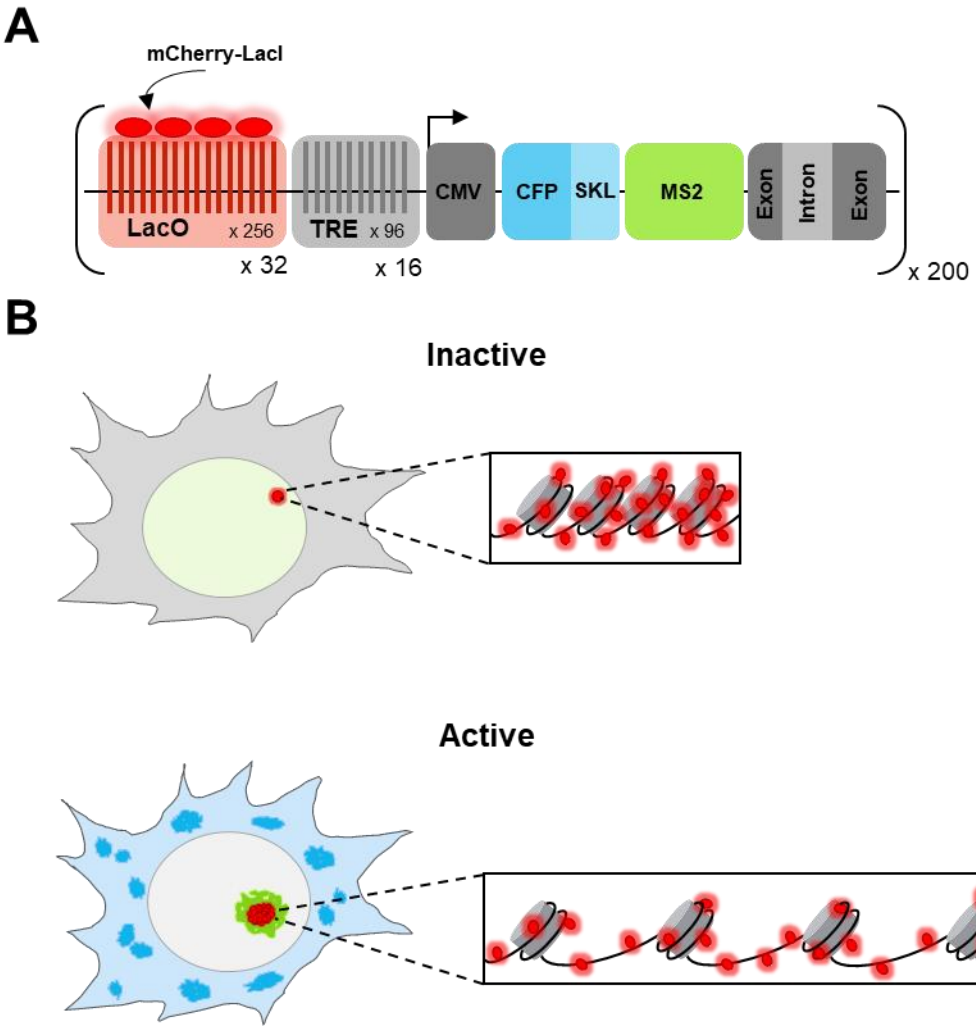
In the literature, the U2OS 2-6-3 reporter gene array has been used extensively to visualize the dynamics of gene expression. The reporter cell line has been utilized in

mapping the kinetics of co-transcriptional processing (Brody et al., 2011; Hochberg-Laufer et al., 2019; Varia et al., 2013), liquid-liquid phase separation (Boija et al., 2018; Chong et al., 2018; Schneider et al., 2021), mitosis (Zhao et al., 2011), and nuclear organization (Giri et al., 2015; Tripathi et al., 2012; Youmans et al., 2018) among others. LacI or TetON fusion proteins allow molecular tethering of factors to the reporter gene locus (Giri et al., 2015, Tripathi et al., 2012). Altogether, U2OS 2-6-3 cell line provides a way to indirectly visualize DNA, RNA, and protein at an inducible gene locus by microscopy.

<u>Figure iii:</u> Cartoon of the mU2OS reporter gene array

A. Adapted from Janicki et al. 2004 (not to scale). The U2OS 2-6-3 reporter gene array consists of approximately 200 copies of the p3216PECMS2β plasmid. Each plasmid copy is 19.19 kilobases in length, and the whole array is stably integrated into chromosome 1p36. The plasmid consists of 5' Lac operator repeats (bp 34 – 9383), TRE (bp 9398 – 13475) and CMV minimal promoter (bp 13486 – 13613). The transcription unit (bp 13539 – 16617) consists of a CFP-SKL fluorescent peroxisomal targeting sequence (bp 13646 – 14464), MS2 module (bp 14497 – 15809), and β-globin unit (exon-intron-exon-polyA signal-polyA site-3' UTR; bp 15811 – 16974). Downstream of the transcription unit is the pUC-derived vector ori and amp sequences (bp 17128 – 19186). In mU2OS cells, mCherry-LacI is stably incorporated and binds to the Lac operator repeats.

B. **Inactive.** The mU2OS reporter gene array is observed by stable mCherry-LacI expression as a bright, rounded dot by microscopy when transcriptionally silent (red). Expression of YFP-MS2BP does not label the inactive locus because reporter transcript is not present (green). Reporter protein product CFP-SKL is not produced. Reporter gene chromatin is highly condensed and heterochromatic (inset). **Active.** The mU2OS reporter gene array is observed by stable mCherry-LacI as an irregularly shaped, globular, enlarged locus by microscopy when transcriptionally active (red). YFP-MS2BP labeled MS2 stem loops accumulate at the active locus to visualize reporter transcripts (green). Reporter protein product CFP-SKL is translated and localized to peroxisomes (blue). Reporter gene chromatin is fully decondensed (inset). Not to scale.


A
mCherry-LacI
LacO
x 256
x 32
TRE
x 96
x 16
CMV
CFP
SKL
MS2
Exon
Intron
Exon
x 200
B
Inactive
Active

CHAPTER II

OBJECTIVES

Expression of protein-coding genes is a complex operation requiring tight spatial and temporal coordination of co-transcriptional pre-mRNA processing factors. Although individual steps of transcription are relatively well understood, not much is known about the transitions governing promoter activation versus inactivation, the chromatin modifications driving these transitions, or the dynamic coupling of transcription to downstream processing steps. Moreover, gene-specific chromatin modifications driving transcription regulation are not well known. The long-term goal of this project is to understand the relationship between SON-dependent chromatin changes and gene regulation.

SON has recently been identified at gene promoters, revealing a new function for SON in gene regulation via chromatin association. Combined with our preliminary findings that SON localizes to a transcriptionally inactive reporter gene locus, we sought to determine the dynamics of SON's association with chromatin and its effect on gene expression. In my thesis, I utilize the U2OS 2-6-3 reporter gene array to elucidate how SON associates with a heterochromatic reporter gene locus, and the role of SON in gene regulation via chromatin association. **My overall hypothesis is that SON chromatin association regulates gene expression.**

Aim 1: D**efine how SON interacts with the transcriptionally inactive U2OS 2-6-3 reporter gene locus**.

In aim 1, I characterized SON localization to the U2OS 2-6-3 reporter gene locus. Colocalization studies were conducted to assess SON's localization with the inactive vs active reporter gene locus, and whether SON association with the locus was reversible and or/correlated with transcription status. I also identified the protein domains required for reporter locus colocalization, and where along the reporter gene SON associates.

Aim 2: Identify how SON regulates gene expression through chromatin association.

In aim 2, I determined how SON affects chromatin condensation at the reporter gene locus. SON was depleted from reporter cells and locus organization observed. Next, the effects of RNA pol II inhibition on SON-depleted chromatin organization were assessed to see if SON's chromatin organization function was dependent on transcription elongation. I also determined how SON depletion affects the localization of epigenetic markers. Finally, SON's role in genome-wide chromatin accessibility was examined.

CHAPTER III

MATERIALS AND METHODS

Cell Lines and Plasmids

Human osteosarcoma U2OS 2-6-3 cells were a gift from David Spector (Cold Spring Harbor Laboratory; Janicki et al., 2004). U2OS 2-6-3 cells stably expressing mCherry-LacI (mU2OS) were a gift from Prasanth Kannanganattu (University of Illinois at Urbana-Champaign). CFP-LacI, pEYFP-MS2BP, pTetON, and pTetOFF were a gift from David Spector (Cold Spring Harbor Laboratory; Janicki et al., 2004). siRNA-refractory YFP-SON plasmids were previously developed by Alok Sharma (Sharma et al., 2011).

Cell Culture

U2OS-263 cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% Tet. approved fetal bovine serum (R&D Systems) and 1% penicillin/streptomycin. Cells were incubated at 37° C and 5% CO_2. Cells were kept under passage 25 and maintained at greater than 30% confluency to preserve reporter gene integrity.

Antibodies

Affinity-purified rabbit polyclonal antibodies were generated at Covance (Denver, PA) against the C-terminal peptide sequence SPNKKHAKATAATV of SON (1:500; WU13) and C-terminal peptide sequence ERPSTTKDKHKEEDKNS of Btf (1:1000; WU10). TRAP150 antibody was purchased from Bethyl (1:1000; Cat#: A300-956A). The following additional antibodies were purchased from Abcam. H3K9me3 (ab8898), H3K4me3 (ab8580), H3K9ac (ab4441), HP1a (ab77256). Anti-RNA pol II was purchased from Diagenode (Cat#: C15200004-10)

Primer Design

Primers along the 2-6-3 transcript first exon, intron, intron-exon boundary, and CFP open reading frame were designed using IDT PrimerQuest by Jacob Ward. SON and GAPDH primers were previously designed by Vishnu Battini. Primer sets along the U2OS 2-6-3 gene (A, B, C, D, E) replicated those used in Janicki et al. (2004), gene desert negative control gene (OR4F5), and active promoter positive control genes (ACTB-2; Pol2-69) were a gift from Joanna Wysocka (Stanford University). The KAT5 primer set was designed using IDT PrimerQuest against the SON ChIP-seq peak (from Jennifer Pence). A list of primer sequences is available in Table 1.

Table 1: List of Primer Sequences

Primer Set	Sequence
263 ECFP Left 13710-13729	ACGTAAACGGCCACAAGTTC
263 ECFP Right 13877-13896	AAGTCGTGCTGCTTCATGTG
263 Beta Intron Left 15933-15952	TGTATCACCATGGACCCTCA
263 Beta Intron Right 16122-16141	ACCCTGATTGCCTTGAAAAA
263 Beta Exon Left 16499-15518	ATGCCCTGGCTCACAAATAC
263 Beta exon Right 16647-16666	CCCATATGTCCTTCCGAGTG
263 Beta Junction Left 16371-16390	TTCATGCCTTCTTCTCTTTCC
263 Beta Junction Right 16469-16488	GCCACCACCTTCTGATAGGC
SON Forw RT	CAGAACTACGATATAAGCC
SON Rev RT	GATACAACTGACAGTTCTGC
GAPDH Forw RT	ATGGGGAAGGTGAAGGTCGGAG
GAPDH Rev RT	CATGTAGTTGAGGTCAATGAAGG
KAT5 prom peak 2 Forw q	ATACACGTGACCAGCTTCG
KAT5 prom peak 2 Rev q	CGAGCCCTAGGTGGAAAC

qRT-PCR

Total RNA extraction was performed using the Qiagen RNeasy kit (Qiagen; Cat#: 74104) per manufacturer's instructions. RNA concentration was determined by Nanodrop (Thermo Fisher Scientific). cDNA was synthesized using qScript cDNA Synthesis Kit from Quantabio (VWR; Cat#: 101414-098) from 500 ng-1 µg of RNA. RT-PCR was performed in Cepheid Step RT-PCR machine for 40 cycles. cDNA samples were then diluted (1:10 for 1 µg RNA input) and added to a 96-well plate with 2x PerfeCTa SYBER Green SuperMix Low ROX (Quantabio from VWR; Cat#: 101414-158), nuclease free water, and 2 µL of 5 µM primer mix per well for a 20 µL reaction volume. qPCR was performed for 40 cycles in the QuantStudio Flex 7. Fold change was calculated using the ΔΔCt method (Pfaffl, 2001) and normalized to GAPDH.

Transfection

U2OS 2-6-3 cells with stably expressed mCherry-LacI were transfected as previously described (Janicki et al., 2004) by electroporation using a Gene-Pulser II (Bio-Rad) (170 V, 950 µF) and 40 µg sheared salmon sperm DNA as a carrier. Cells were transfected with 2 µg of plasmid (rtTA (pTetON); CFP-LacI; YFP-MS2BP), and 3 hours after transfection the reporter gene was activated by addition of doxycycline (1 µg/mL) to the culture medium. Cells were processed for microscopy 2.5 hours after doxycycline addition. mU2OS promoter inactivation was achieved by electroporation of pTetOFF. Three hours post-transfection, doxycycline was added to the culture medium and cells processed for microscopy 2.5 hrs after dox addition. For time course experiments, dox

addition began 3 hours after pTetON transfection until all cells were harvested 5.5 hours post transfection. For dox washout, cells were treated with media containing dox (1 µg/mL) for 2.5 hours, followed by two washes with 1x PBS and replenishment with media without dox for 3 hours before immunofluorescence processing.

Coverslip Preparation

Coverslips were prepared in large batches by washing thoroughly in one part nitric acid and two parts hydrochloric acid for 2 hours with occasional stirring. They were then rinsed in distilled water until the wash reached pH 7.0 and were subsequently stored in 70% ethanol. Coverslips were passed through a Bunsen burner flame for sterilization before experimental use.

Immunofluorescence

Indirect immunofluorescence was performed as described (Sharma et al. 2010). Briefly, cells were grown on acid-washed glass coverslips in a 6-well dish. Cells were washed with phosphate buffered saline (PBS) and immediately fixed with 2% formaldehyde in PBS for 15 min at room temperature. After three PBS washes (5 min each), cells were permeabilized with 0.2% Triton X-100 (in PBS) for 5 min at room temperature. The cells were rinsed 3 times for 5 min each in PBS containing 0.5% normal goat serum (NGS) and incubated with primary antibody for 1 hour at room temperature in a humidified chamber. Goat anti-HP1a was blocked in PBS + BSA instead of NGS. Primary antibody used was anti-SON WU13 (1:500) for all experiments. After three

washes with PBS/NGS, cells were incubated with secondary antibody goat anti-rabbit conjugated with Alexa Fluor 488 (1:500; Jackson Immunoresearch Cat # 711-545-152). Mouse antibodies were incubated with secondary antibody goat anti-mouse FITC (1:500; Jackson Immunoresearch). The cells were washed twice in PBS, and a final wash of DAPI (1:20,000) in PBS was used to label DNA. Coverslips were then mounted on slides with Vectashield antifade mounting medium (Vector labs Cat # H-1000) and stored at -80° C until imaged.

Chromatin Immunoprecipitation

Protocol was adapted from Agilent. mU2OS cells at 80-90% confluency (~$1x10^7$ cells/dish) were fixed by adding 1 mL formaldehyde solution (50 mM HEPES-KOH, pH 7.5; 100 mM NaCl; 1 mM EDTA; 11% formaldehyde) directly to the culture medium and incubating at room temperature for 10 min. Fixation was stopped by adding 500 µL of 2.5 M glycine to quench the formaldehyde and incubated at room temperature for 5 min. Cells were washed twice with 1x PBS and harvested using a rubber cell scraper (USA Scientific; Cat#: CC7600-0220) before transferring to a conical tube on ice and centrifuged to pellet. Cells were resuspended in 1X PBS to clear the residual culture medium, centrifuged, the supernatant discarded, and the cell pellet flash frozen at -80° C for 30 min.

The remaining steps were done on ice unless indicated. The frozen cell pellet was then resuspended in cold Lysis Buffer 1 (50 mM HEPES-KOH, pH 7.5; 140 mM NaCl; 1 mM EDTA; 10% glycerol; 0.5% NP-40; 0.25% Triton X-100; 1X Roche cOmplete

ULTRA protease inhibitor tablet EDTA-free Cat#:5892953001) at a concentration of 1.66 mL per 1 x 10^7 cells and rotated at 4° C for 10 min, followed by 5 min centrifugation at 1,350 x g and the supernatant discarded. The pellet was then resuspended in Lysis Buffer 2 (10 mM Tris-HCl, pH 8.0; 200 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 1x protease inhibitor tablet) at a concentration of 1.66 mL per 1 x 10^7 cells and rotated at room temperature for 10 min, followed by 5 min centrifugation at 1,350 x g and the supernatant discarded. Pelleted nuclei were then resuspended in Lysis Buffer 3 (10 mM Tris-HCl, pH 8.0; 100 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 0.1% Na-Deoxycholate; 0.5% N-lauroylsarcosine; 1X protease inhibitor tablet) at a concentration of 1 mL per 1 x 10^7 cells. The lysate was then sonicated for 2 x 7.5 min on HIGH 30 sec on/off cycles (Diagenode Bioruptor), 1/10th volume of 10% Triton X-100 added and centrifuged at 20,000 x g for 10 min to pellet cellular debris. The freshly prepared lysate was collected and divided into 440 µL extract/ChIP reaction. Each ChIP had 5 µL target antibody or no antibody control, and reactions rotated overnight at 4° C.

Protein A magnetic Dynabeads (Invitrogen; Cat#: 10002D) were prepared by washing 3 x 1 mL Wash Solution (1X PBS; 0.5% BSA w/v) and resuspended in Lysis Buffer 3 for equilibration. Bead volume 100 µL/reaction was added to the protein-antibody complexes and rotated for 4 hours at 4° C. Immunoprecipitated protein complexes were captured by a magnetic stand on ice and washed 5 x 1 mL in cold RIPA Wash Buffer (50 mM HEPES-KOH, pH 7.5; 250 mM LiCl; 1 mM EDTA; 1% NP-40; 0.7% Na-Deoxycholate) followed by a final wash in 1 mM TE + 50 mM NaCl. Beads were centrifuged for 3 min at 960 x g, the supernatant removed, and 210 µL Elution Buffer (50 mM Tris-HCl, pH 8.0; 10 mM EDTA; 1% SDS) added to each ChIP.

Complexes were eluted at 65° C for 15 min, centrifuged 1 min at 16,000 x g, and 200 µL eluate transferred to a fresh tube. Eluates and WCE were adjusted to 50 mM NaCl by adding 2 µL 5 M NaCl to each tube, and all samples were reverse crosslinked overnight at 65° C.

Reverse crosslinked samples were diluted with 200 µL TE and digested with 8 µL of 10 mg/mL RNAse A for 30 min at 37° C followed by 7 µL $CaCl_2$ Stock Solution (300 mM $CaCl_2$ in 10 mM Tris HCl, pH 8.0) addition and 4 µL of 20 mg/mL proteinase K digestion for 30 min at 55° C.

DNA was extracted by adding 400 µL phenol:chloroform:isoamyl alcohol (Invitrogen; Cat#: 15593-031), vortexed, and centrifuged for 5 min at 16,000 x g. The supernatant was transferred to a fresh tube and 16 µL of 5 M NaCl, 1.5 µL of 20 µg/µL glycogen (Roche; Cat#: 10-901-393-001), and 880 µL ice cold 100% ethanol added before the mixture was cooled for 45 min at -80° C. The cold mixture was then centrifuged for 10 min at 20,000 x g at 4° C to create DNA pellets, the supernatant removed, and pellets left to dry for at least 15 min before purified DNA was resuspended in 200 µL nuclease-free water and qPCR performed.

Input WCE ChIP-DNA in nuclease-free water was diluted 1:10 compared to ChIP samples, resulting in 1.14% input in qPCR reactions. Each qPCR well contained 10 µL 2x SYBER SuperMix Low ROX, 4 µL ChIP DNA, 4 µL nuclease-free water, and 2 µL of 5 µM primer mix. After qPCR in the QuantStudio Flex 7, data was analyzed via the percent input method (Thermo Fisher). Input Cts were adjusted to 100% by $Ct_{Input} - 6.459$ (dilution factor of 1.14% input = $\log_2 88$). Then, % input calculated by 100 x $2^{(\text{adjusted input} - \text{Ct sample})}$.

Silencing and Overexpression

mU2OS cells were seeded in a 6-well dish prior to transfection. A double dose of Oligofectamine (Invitrogen from Fisher Scientific; Cat#: 12-252-011) transfection was performed (at 24 hours and 48 hours) with 60 pmol of either non-targeting control siRNA duplex for luciferase (Dharmacon siGENOME; Cat#: D-001206-14-20) or SON siRNA duplexes 1 (Dharmacon siGENOME; Cat#: D-012983-01-0050) and 4 (Dharmacon siGENOME; Cat#: D-012983-05-0050). 72 hours post-transfection, cells were electroporated with carrier DNA or pTetON plasmids as described in transfection methods. For overexpression, YFP-siR-SON plasmid was transfected into cells via Lipofectamine 3000 (Invitrogen from Fisher Scientific; Cat#: L300008) 24 hrs post-seeding per manufacturer's instructions and processed for microscopy 48 hrs post-transfection.

DNAse-qPCR

Protocol was adapted from Nepon-Sixt and Alexandrow (2019). mU2OS cells were treated with non-targeting siRNAs or SON si1+4 as described in silencing methods above in 5 wells each of a 6-well dish. Cells were washed twice with PBS, pooled by control vs SON KD condition, and collected by rubber policeman scraping before resuspension in 300 µL Chromatin Lysis Buffer and 5 min incubation on ice. Whole cell lysate was then centrifuged and pelleted nuclei resuspended in Chromatin Buffer. Samples were divided into 3 tubes x 100 µL representing "uncut" and "cut," and third tube to save. "Cut" samples had 3 µL DNase I (Thermo Fisher Scientific; Cat #:

FEREN0521) added and both samples had 10 µL $MgCL_2/CaCl_2$ (Thermo Fisher Scientific; Cat#: FERB43) added before subjection to 5 min at 65° C. The DNase digestion was stopped with 0.25 M EDTA and 100 mM EGTA and incubated at 65° C for 10 min. SDS (10%) was added and reactions were sonicated for 8 min on high for 30 sec on/off cycles (Diagenode Bioruptor). After sonication, RNase A (Thermo Scientific; Cat#: FEREN0531) was added and reactions incubated at 37° C for 30 min, followed by overnight proteinase K (Thermo Scientific Cat#: FEREO0491) digestion at 42° C. The next day, DNA was extracted using PCIA and purified DNA was analyzed by 1.5% agarose gel electrophoresis and qPCR. Data was converted as percent of "cut" chromatin relative to "uncut" chromatin by calculating $100 / 2^{(\text{cut Ct-uncut Ct})}$ for control and SON KD cells. Then, SON KD percent input chromatin was normalized to Control cells to determine the amount of chromatin left after DNase treatment in SON KD cells compared to the amount of chromatin left after DNase treatment of Control cells. A final value of 0.6 for SON KD cells, for example, would indicate that DNase digestion of SON KD cells resulted in qPCR product recovery of 60% compared to Control cells.

α-Amanitin Treatment

For inhibition or RNA Polymerase II transcription, cells were electroporated and plated on coverslips. After 3 hours, the medium was changed with media containing 50 µg/mL α-amanitin. Dox was added (1 µg/mL) after 3 hours amanitin treatment, and cells were processed for immunofluorescence after 3 hours dox for a total α-amanitin treatment of 6 hours.

Microscopy

Images were acquired on Olympus IX83 inverted fluorescence microscope using a 100x objective (1.3 numerical aperture; Olympus) and low autofluorescence immersion oil (Olympus, n = 1.518; Fisher Sci Cat #NC0297589). Exposure time was based on optimal signal-to-noise ratio for each image. Raw images were saved as TIFs, imported into ImageJ, pseudo-colored, and saved as individual channel JPEGs or PNGs. Images were cropped and stretched as indicated by scale bar (5 μm). Some images underwent background subtraction in ImageJ to reduce background fluorescence and those are denoted in the associated figure legends.

Cell Scoring

Raw TIF images acquired on IX83 microscope were analyzed for colocalization using Olympus CellSens Dimension software. The region-of-interest (ROI) was determined by manually drawing a 2-point circle, oval or freehand polygon around the mCherry-LacI labeled U2OS 2-6-3 gene locus. To assess colocalization between mCherry-LacI and FITC-SON, a red vs. green pixel scatter plot was generated for the identified locus and Pearson Correlation Coefficient (PCC) generated using the Colocalization function in CellSens. PCCs were designated as negative colocalization ($PCC < -0.1$), no colocalization ($-0.099 < PCC > 0.099$), or positive colocalization ($PCC > 0.1$). I considered a $PCC \geq 0.1$ to indicate colocalization because the heterochromatic, highly condensed inactive reporter locus can limit antibody accessibility and make it difficult to assess direct colocalization by microscopy. Also, the spatial separation

between the Lac Operator repeats and gene body lowers the PCC of regulatory factors that colocalize with the non-labelled downstream reporter gene (Varia et al., 2013). For epigenetic factor localization in aim 2b, loci were designated as positive colocalization (PCC > 0.2) because the diffuse nuclear staining of these factors can artificially inflate PCCs, so a higher threshold for colocalization was used. When comparing locus size, cells were scored based on PCC as well as ROI area. Loci < 2.0μm^2 were scored small/closed/inactive and loci ≥ 2.0μm^2 scored large/open/active.

Chromatin Compaction DAPI Analysis

DAPI-stained nuclei from the same α-amanitin experimental slides described in Figures 16-18 were imaged under the 100x objective and individually cropped to yield one cell per 8-bit TIFF image. For each condition, 20-45 individual cells were captured (separate cells imaged than those analyzed in the SON α-amanitin Figures 16-18). Image processing and numerical CCP data was obtained in collaboration with undergraduate researcher Joshua Ward with MATLAB code adapted from Irianto et al. (2014). I conducted the experiment, imaged the cells, analyzed the resulting numerical data, and created the figures.

A summary of the MATLAB workflow from raw image to CCP calculation is as follows. First, a given cell image was read. Threshmode function was used to generate a pixel intensity histogram for each cell, identify the peak pixel intensity, and identify the pixel threshold separating cell boundary from background (threshold was empirically determined, Figure 27). Cells underwent two iterations of smoothing using the fspecial

function, signal threshold was applied to each smooth image, and holes in the image were filled with a black background. The nucleus from the raw input image was placed on a black background and pixel intensities redistributed to make the range of intensities 0 to 255. Images were reduced by a pixel reduction factor of 4 (empirically determined, Figure 27). Sobel edge detection was applied using the "edge" function and holes were filled with black background. The cell perimeter of the Sobel edge detected nucleus was removed and the inner nucleus was placed on a background. Nucleus area was determined from the cell perimeter. The length of the Sobel edge vector was identified and the CCP was calculated by taking the number of pixels identified in the Sobel edge detection vector divided by the total pixel area of the nucleus. Individual CCPs were calculated for each image in a given condition. The full MATLAB code is available in Appendix B.

Statistical analysis

The number of biological replicates for a given data set is described in figure legends. When multiple replicates of cell scoring data was collected, the data from all replicates was pooled for a given condition for statistical analysis. In microscopy studies, the experimental unit consisted of all cellular conditions that were tested in tandem originating from the same seeding cell population. When only selected conditions from a given experimental unit are displayed in a figure, statistical analysis was performed on the entire data set. For studies comparing the average PCC between different transcription conditions, an unpaired two-tailed Student's T-test was performed with Welch's correction without assumption of equal standard deviation. For studies

comparing transcription status and SON KD vs. Control, a two-way ANOVA with Sidak's multiple comparisons test was performed. For Figures 28-30 examining transcription status, α-amanitin treatment, and SON KD vs. Control, a three-way ANOVA was performed, followed by Tukey's multiple comparison test. The statistical analyses described above was performed by me using GraphPad Prism version 9.3.1 for Windows, GraphPad Software, San Diego, California USA, www.graphpad.com.

The following tests were conducted using SAS version 9.4 (SAS Institute, Inc., Cary, NC) in collaboration with Michael Bottomley of the Statistical Consulting Center. For Figure 1, independent samples t-test was performed with Satterthwaite's method of calculating degrees of freedom to overcome the violation of constant variance. In Figure 6, a one-way ANOVA and post hoc Tukey's multiple comparisons procedure was employed. For Figures 16-18 examining transcription status, α-amanitin treatment, and SON KD vs. Control, a three-way ANOVA with Games-Howell procedure for unequal variances was conducted, followed by Tukey's multiple comparison test.

CHAPTER IV

AIM 1 RESULTS

Subaim 1a: Characterize SON's localization to the U2OS 2-6-3 locus.

SON is a large, repetitive nuclear speckle protein with established roles in pre-mRNA processing and nuclear speckle organization. SON depletion alters the splicing of pre-mRNA transcripts but does not affect global mRNA levels (Sharma et al., 2011). The specifics of how SON interacts with a gene, where SON associates along a gene, and at what step in gene expression SON is involved is currently unknown. In this specific aim, I investigated the dynamics of SON's role in transcription using the U2OS 2-6-3 reporter gene array. The U2OS 2-6-3 reporter gene locus allows us to visually detect locus chromatin organization, transcription and pre-mRNA processing status, mRNA export, and reporter protein production. I used colocalization analyses to assess SON localization to the reporter locus following transcription activation/inactivation to identify how SON interacts with the reporter array.

SON colocalizes with the transcriptionally inactive U2OS 2-6-3 locus

The U2OS 2-6-3 reporter gene locus allows visual detection of locus chromatin organization, transcription and pre-mRNA processing status, mRNA export, and reporter

protein production (Janicki et al., 2004; schematic in Figure 1a). An initial pilot study was performed to test the visibility of the mU2OS locus, the efficiency of pTetON transfection, and the selectivity of reporter gene activation with only pTetON in the presence of doxycycline (results in Appendix A Figure A1-A4). The U2OS 2-6-3 reporter gene is tightly controlled by a tetracycline response element. Reporter gene activation requires expression of rtTA (rTetR fused to the VP16 activation domain) that binds to tetracycline response elements and drives CMV minimal promoter activation and reporter gene transcription only in the presence of doxycycline.

SON localization in relation to the mU2OS inducible reporter locus was assessed to determine the role of SON in gene expression. SON colocalization with mCherry-LacI was determined by Pearson Correlation Coefficient (PCC), which is a measure of the degree of correlation between two variables generated from a FITC vs. mCherry-LacI scatterplot of pixel intensity. A PCC maximum of 1.0 indicates a perfect positive correlation, and a PCC minimum -1.0 indicates a perfect negative correlation. A PCC of 0 indicates no correlation between the two channels.

Unlike pre-mRNA processing factors Btf, TRAP150, and SRSF1 which are observed at the active U2OS 2-6-3 reporter locus (Janicki et al., 2004; Varia et al., 2013), endogenous SON did not colocalize with the active locus (Figure 1b; panels a-c) in the majority of large loci. For the 103 total cells scored in the “TetON + dox” condition, loci had an average PCC of 0.014 ± 0.180 (Figure 1c). Approximately 20% of the loci scored remained condensed after 2.5 hrs post-dox, most likely because of TetON transfection efficiency or some cells not being activatable in the experimental timeframe.

Interestingly, SON instead colocalized with the inactive, condensed reporter locus (Fig. 1b; panels d-f). SON colocalization with mCherry-LacI was observed in the majority of inactive cells with small loci (locus area < 2.0 μm^2). SON-LacI average PCC was significantly increased in the inactive condition compared to the transcriptionally active condition (Figure 1c). For 101 total cells scored, loci had an average PCC of 0.133 ± 0.249. Association of SON with the transcriptionally inactive gene locus suggests a previously unknown role for SON in gene silencing.

To rule out that SON colocalization could be due to some 'leaky transcription' of the reporter gene causing pre-mRNA processing factor recruitment, SON colocalization with mCherry-LacI was assessed in a series of transcription controls (TetON schematic in Figure 2a). Wild-type loci (non-transfected control; Figure 2b; panels a-c), mock transfected with carrier DNA only (electroporation control; Figure 2b; panels d-f), loci treated with "dox only" (to assesses any spontaneous locus activation by doxycycline only; Figure 2b; panels g-i), or loci treated with "TetON only" to examine loci activated by TetON transfection alone (Figure 2b; panels j-l). In all cases except "TetON + dox" SON colocalized with the mCherry-LacI labeled locus in the majority of cells scored (Figure 2c). These results reaffirm the initial cell line characterization experiments (Appendix A) that the mU2OS locus is tightly regulated and only cells expressing TetON in the presence of dox can be transcriptionally active. SON colocalizes with the transcriptionally silent locus in all negative controls and has not been observed to colocalize with the transcriptionally active mU2OS locus.

SON localization was also assessed in U2OS 2-6-3 cells that did not stably express mCherry-LacI. Endogenous SON colocalized with a transcriptionally inactive

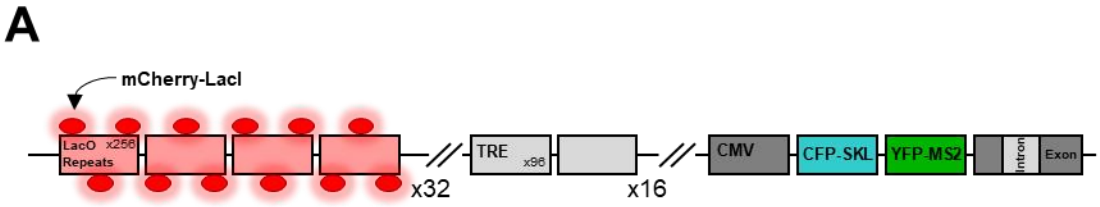
CFP-lacI labelled reporter locus (Figure 3; panels a-c) and did not colocalize with the transcriptionally active CFP-LacI locus (Figure 3; panels f-h). Transcription status was assessed by YFP-MS2BP colocalization with the active gene locus (Figure 3; panels d, i). These results, taken together with the mU2OS SON colocalization studies in Figure 1 and 2, indicate endogenous SON localization with the 2-6-3 gene is specific to transcriptionally silent loci, and SON is not detectable at the transcriptionally active, decondensed locus.

<u>Figure 1:</u> SON colocalizes with the transcriptionally inactive U2OS 2-6-3 gene locus.

A. Schematic representation of the reporter construct used to generate U2OS 2-6-3 cell line (adapted from Janicki et al., 2004).

B. Endogenous SON is not detected at activated mU2OS reporter loci (a-c; PCC < 0.1; 67/102 cells) but is colocalized with inactive loci (d-f; PCC ≥ 0.1; 58/100 cells) Upper right value in LacI channel indicates locus area (μm^2); upper right value in merge channel indicates PCC. Scale bar = 5 μm.

C. Plot of pooled locus PCCs from n = 2 independent replicates with 101 (inactive) or 103 (active) total cells scored. Bars denote mean and standard deviation. P-value = 0.0001 and was computed by independent samples t-test with Satterthwaite's correction for unequal variance.

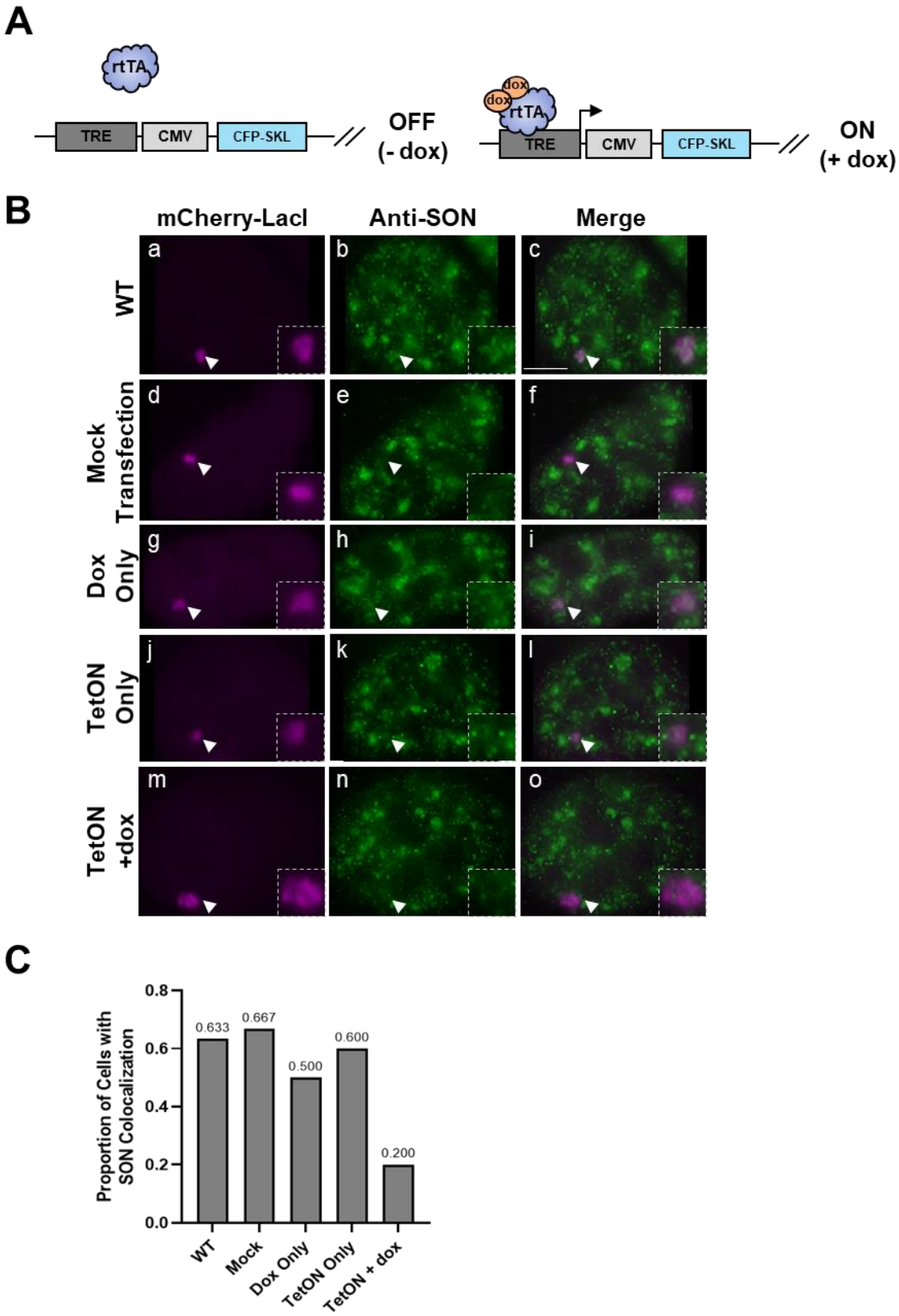

A
mCherry-LacI
LacO x256 Repeats
x32
TRE x96
x16
CMV
CFP-SKL
YFP-MS2
Intron
Exon
B
mCherry-LacI
SON
Merge
Active
Inactive
a
5.11
b
c
-0.036
d
1.22
e
f
0.429
C
***
1.0
0.5
0.0
-0.5
-1.0
PCC
Inactive
Active

<u>Figure 2:</u> SON localization at the U2OS 2-6-3 locus is not due to leaky reporter transcription.

A. Schematic representation of the TetON system. When pTetON is expressed in mU2OS in cells without doxycycline, the encoded rtTA cannot bind to the TREs and the reporter gene remains inactive. In the presence of doxycycline, rtTA binds to the TREs, activating reporter gene expression from the CMV promoter.

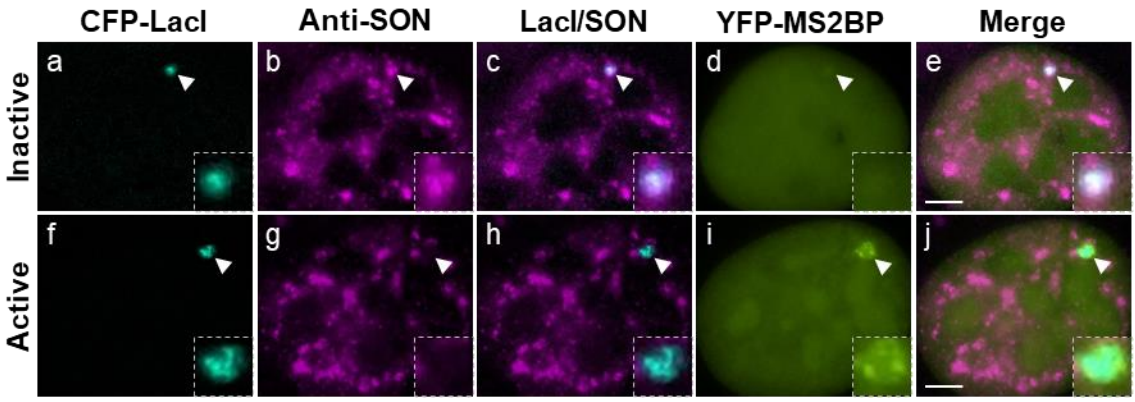
B. U2OS 2-6-3 cells stably expressing mCherry-lacI were either left untransfected and uninduced (WT; a-c), transfected with carrier DNA and uninduced (mock; d-f), or induced with dox (dox only; g-i), or transfected with TetON and uninduced (TetON only; j-l) or induced with dox (TetON+dox; m-o). At 2.5 h after dox addition, the cells were processed for immunofluorescence localization with anti-SON WU13. Scale bar = 5 μm.

C. Chart shows the percentage of cells showing SON colocalization with mCherry-LacI (PCC $\geq$ 0.1), n = 30 cells for all conditions from n = 1 biological replicate.


A
rtTA
TRE
CMV
CFP-SKL
OFF
(- dox)
dox
dox
rtTA
TRE
CMV
CFP-SKL
ON
(+ dox)
B
mCherry-LacI
Anti-SON
Merge
WT
Mock Transfection
Dox Only
TetON Only
TetON +dox
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
C
Proportion of Cells with SON Colocalization
0.0
0.2
0.4
0.6
0.8
0.633
0.667
0.500
0.600
0.200
WT
Mock
Dox Only
TetON Only
TetON + dox

<u>Figure 3:</u> SON colocalizes with the inactive U2OS 2-6-3 locus in cells without stably incorporated mCherry-LacI.

Reporter loci were labelled with CFP-LacI and MS2 loops labelled with YFP-MS2BP in inactive (noTetON-dox; a-e) or active (TetON+dox; f-j) cells. 2.5 h post-dox, cells were immunolabeled with anti-SON WU13 and a Texas Red-conjugated secondary antibody. Background subtraction was performed in the CFP and Texas-Red channels in ImageJ to minimize background fluorescence. Scale bar = 5 μm.


CFP-LacI
Anti-SON
LacI/SON
YFP-MS2BP
Merge
Inactive
Active
a
b
c
d
e
f
g
h
i
j

SON is depleted from the mU2OS locus before visible chromatin decondensation.

SON colocalizes with the heterochromatic mU2OS locus and is not detectable at the transcriptionally active locus, but the dynamics of this shift in localization are unknown. Reporter transcript can be visualized as early as 7.5 min post-dox treatment, peaking around 2.5 hrs post-induction when the locus chromatin also becomes fully decondensed (Janicki et al., 2004). mU2OS cells were processed for immunolocalization of SON at 5 min, 10 min, 30 min, 60 min, 90 min, 120 min, or 150 min following locus activation to determine when SON is no longer detectable at the locus (representative images in Figure 4a). At each timepoint, cells were scored for locus area and SON colocalization with mCherry-LacI as Colocalized (PCC $\geq$ 0.1), No Correlation (-0.1 < PCC < 0.1) or Removed (PCC $\leq$ -0.1). The third classifier "Removed" was added because a negative correlation would indicate absence of SON at the locus.

The results of the cell scoring are displayed as a heat map in Figure 4b. The proportion of scored loci with SON colocalization decreased within the first 60 min of locus activation. In loci that remained inactive/small at timepoints after 90 min dox, SON was still colocalized, suggesting these cells did not get activated (most likely related to pTetON transfection efficiency). The highest proportion of scored loci were decondensed by 60 min - 90 min dox treatment, and the large loci corresponded with decreased SON colocalization (PCCs near 0).

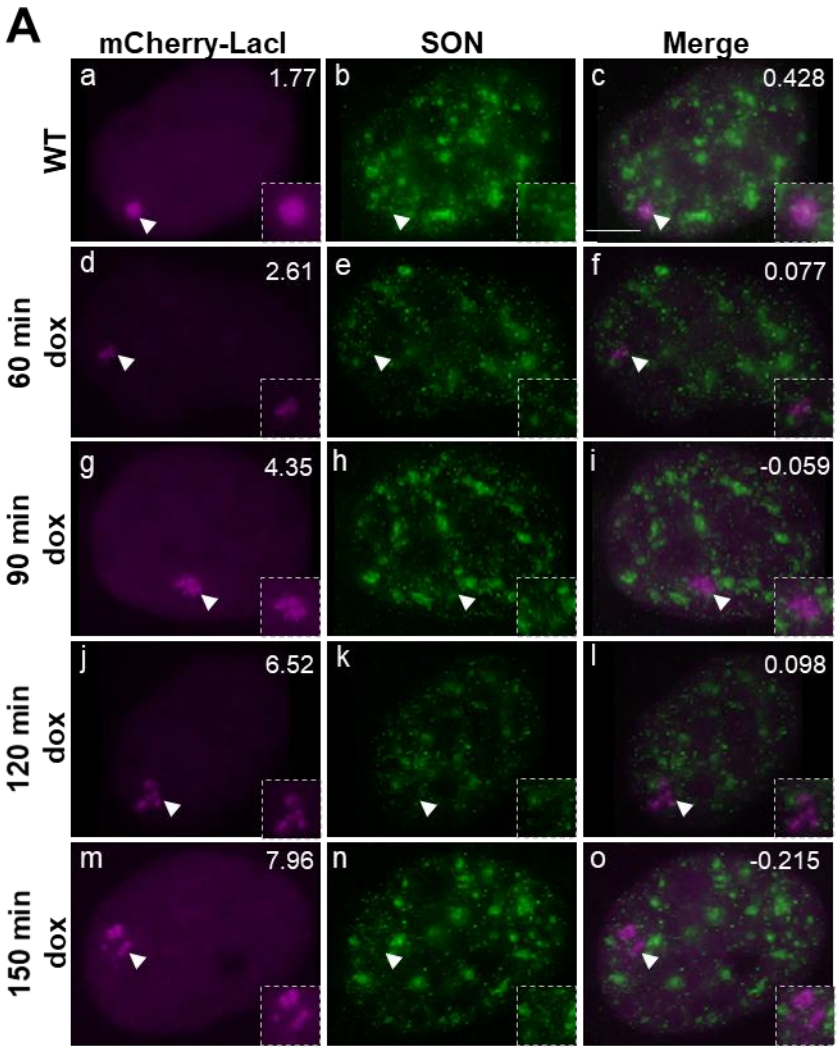
In the 5 min and 10 min post-dox conditions of this experiment, higher order chromatin decondensation of the reporter locus was not detectable, as evidenced by the large number of loci < 2.0 μm^2 (representative images in Figure 5a). A heat map of the early timepoint scoring data is displayed in Figure 5b. Strikingly, at 10 min post-dox,

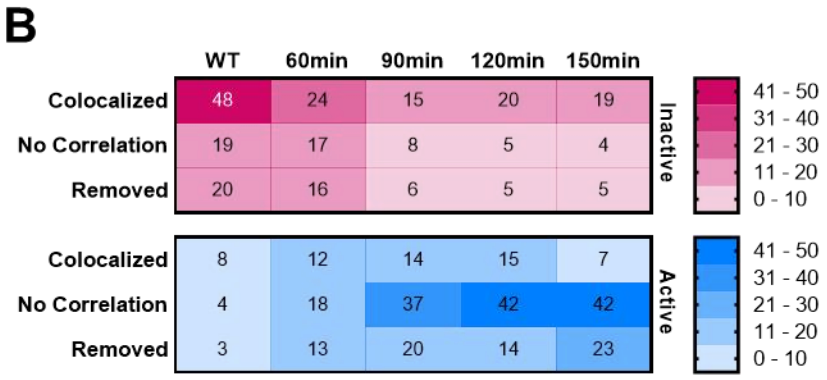
while loci are still condensed, SON was not detectable on the largest proportion of loci scored as indicated by a more negative average PCC (average PCC at 10 min -0.008 ± 0.223). SON depleted reporter loci were not common in the mock transfection condition (mock transfected average PCC 0.162 ± 0.276). At 30 min dox, loci were mostly small and loci that were already decondensed exhibited no SON colocalization. These data indicate that SON is absent from the gene locus before significant locus chromatin decondensation, and before reporter transcript is first detectable at the locus via microscopy.

<u>Figure 4:</u> Activation of the U2OS 2-6-3 reporter locus leads to rapid loss of SON.

A. Wild-type cells (a-c; n = 101 cells) were left without dox, or pTetON transfected cells (d-0) had dox added to cells for 60 min (d-f; n = 100 cells), 90 min (g-i; n = 100 cells), 120 min (j-l; n = 101 cells), or 150 min (m-o; n = 100 cells). All cells were processed for immunolocalization of SON 5.5 hrs post-transfection. Upper right value in LacI channel indicates locus area (μm^2); upper right value in merge channel indicates PCC. Scale bar = 5 μm.

B. Heat map indicating number of loci scored per condition with SON colocalized with mCherry LacI (PCC $\geq$ 0.1), no correlation (-0.1 < PCC <0.1) or removed (PCC $\leq$ -0.1) relative to locus condensation state. Area of loci in pink are < 2.0 μm^2, while loci in blue are $\geq$ 2.0 μm^2, bin = 10 cells. Data represents PCCs of 50-52 cells per condition and pooled for n = 2 biological replicates.

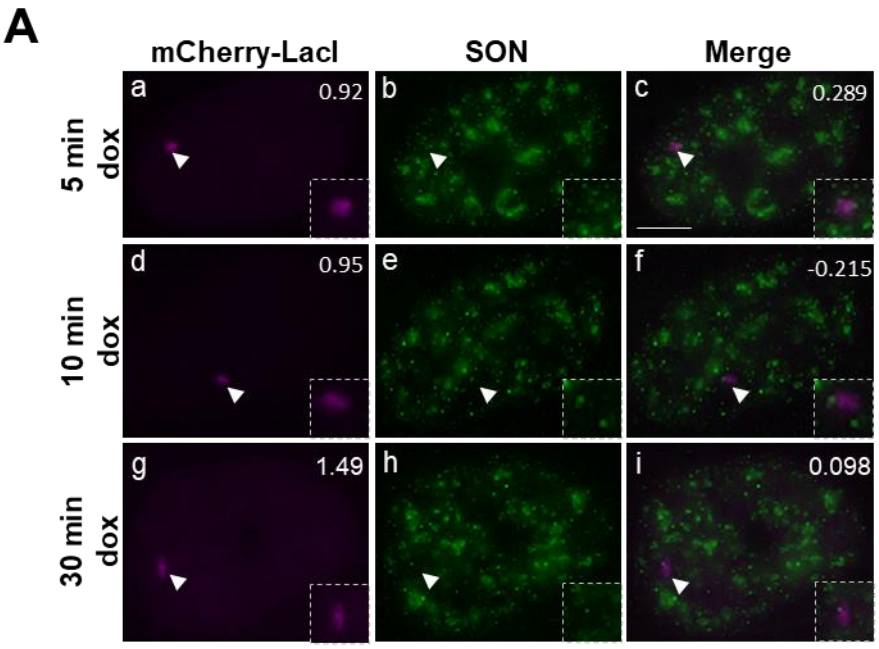

A
mCherry-LacI
SON
Merge
WT
60 min dox
90 min dox
120 min dox
150 min dox
a
1.77
b
c
0.428
d
2.61
e
f
0.077
g
4.35
h
i
-0.059
j
6.52
k
l
0.098
m
7.96
n
o
-0.215

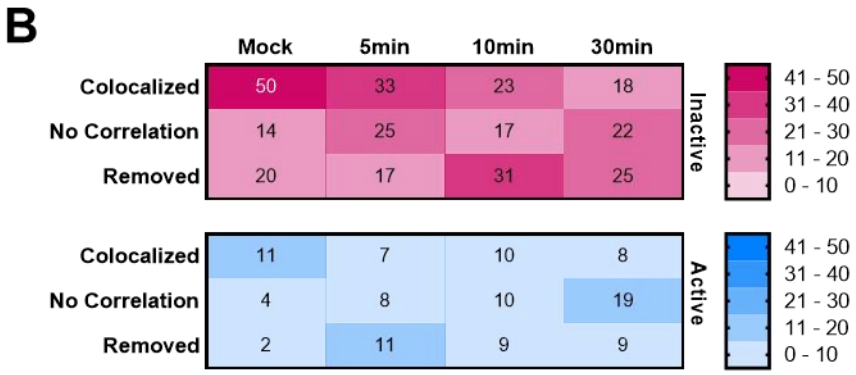

B
WT
60min
90min
120min
150min
Colocalized
No Correlation
Removed
48
24
15
20
19
19
17
8
5
4
20
16
6
5
5
Inactive
41 - 50
31 - 40
21 - 30
11 - 20
0 - 10
Colocalized
No Correlation
Removed
8
12
14
15
7
4
18
37
42
42
3
13
20
14
23
Active
41 - 50
31 - 40
21 - 30
11 - 20
0 - 10

<u>Figure 5:</u> SON is depleted from the activated locus before detectable chromatin decondensation.

A. Expanded experimental data from Figure 4. Post-TetON transfection 3 hrs, dox was added to cells for 5 min (panels a-c; n = 101 cells), 10 min (panels d-f; n = 100 cells), or 30 min (g-i; n = 101 cells). All cells were processed for immunolocalization of SON 5.5 hrs post-transfection. Upper right value in LacI channel indicates locus area (μm^2); upper right value in merge channel indicates PCC. Scale bar = 5 μm.

B. Expanded experimental data from Figure 4. Heat map depicting SON colocalization with mCherry-LacI at early timepoints post-dox treatment when loci are condensed (< 2.0 μm^2). Area of loci in pink are < 2.0 μm^2, while loci in blue are ≥ 2.0 μm^2, bin = 10 cells. Data represents PCCs of 50-51 cells per condition and pooled for n = 2 biological replicates.


A
mCherry-LacI
SON
Merge
5 min dox
10 min dox
30 min dox
a
0.92
b
c
0.289
d
0.95
e
f
-0.215
g
1.49
h
i
0.098


B
Mock
5min
10min
30min
Colocalized
No Correlation
Removed
50
33
23
18
14
25
17
22
20
17
31
25
Inactive
41 - 50
31 - 40
21 - 30
11 - 20
0 - 10
Colocalized
No Correlation
Removed
11
7
10
8
4
8
10
19
2
11
9
9
Active
41 - 50
31 - 40
21 - 30
11 - 20
0 - 10

Endogenous SON colocalization with the mU2OS locus is reversible through reporter gene inactivation.

SON also accumulates at the U2OS 2-6-3 locus following transcription inactivation (Figure 6). Activation of mU2OS cells leads to large-scale chromatin decondensation, and this decondensation is reversible once transcription is turned off by promoter inactivation. Inactivated loci are similar in size to 2-6-3 loci in WT cells where transcription has not been activated (Zhao et al., 2011). Transcription inactivation of fully decondensed loci gives insight into SON's role in gene expression of mU2OS heterochromatic loci by identifying if SON localization at the condensed locus is reversible and/or correlated with transcription status and a condensed chromatin structure.

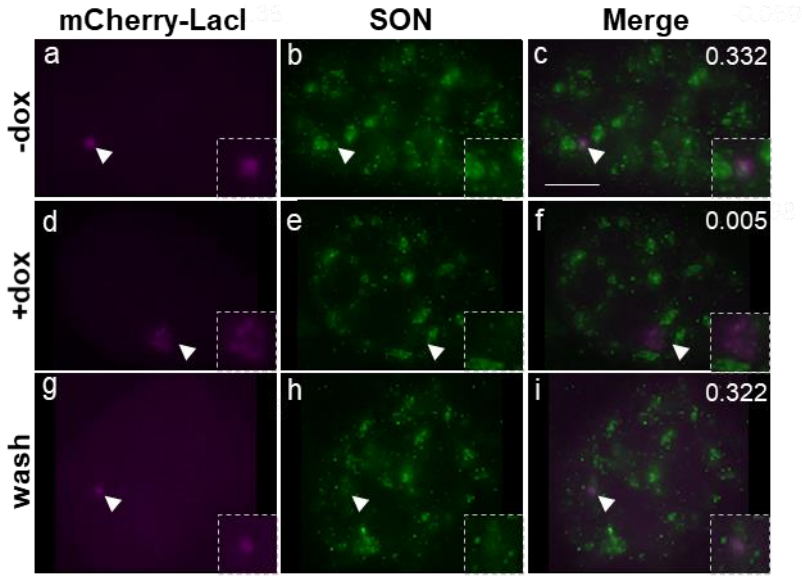
Cells expressing pTetON were either left uninduced (-dox; Figure 6 panels a-c), activated (+dox; Figure 6 panels d-f), or inactivated (wash; Figure 6 panels g-i) that were first induced by addition of dox, followed by removal of dox by media replacement. Activated cells exhibited significantly lower SON-LacI colocalization compared to uninduced cells and wash-inactivated cells (Figure 6;lower). There was no significant difference in SON-LacI colocalization between uninduced and wash-inactivated cells.

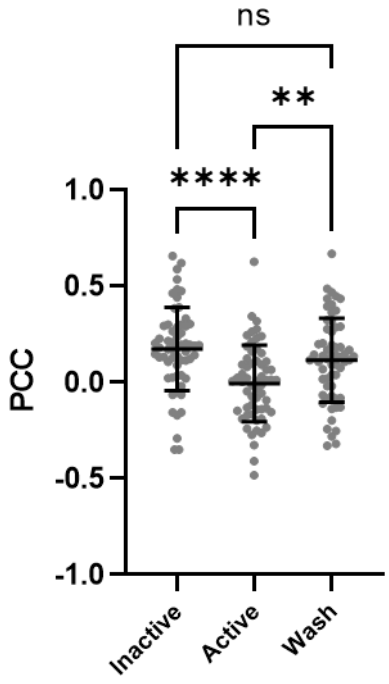
The TetOFF system was also used inactivate the mU2OS locus to assess SON localization change correlated with transcription status. Transfection of cells with pTetOFF encodes tTA, which binds to the TREs promoting CMV promotor activation without the addition of doxycycline. In TetOFF-expressing cells, dox addition inhibits tTA binding to the TREs, inhibiting CMV promoter activation and inactivating reporter gene transcription. A schematic of the TetOFF system is depicted in Figure 7a. SON was

not detectable at the transcriptionally active, decondensed “TetOFF – dox” mU2OS locus (Figure 7b; panels a-c) and was colocalized at the inactivate “TetOFF + dox” condition (Figure 7b; panels d-f). Only about 25% of loci exhibited SON colocalization compared to almost 60% of loci using the media replacement method of transcription activation, perhaps because it takes longer to remove tTA from the promotor to meaningfully inhibit gene expression (TetOFF loci scoring displayed in Figure 7c). Adding dox for a longer time such as 6 hrs may have been more suitable to detect SON at the majority of “TetOFF + dox” inactivated loci.

<u>Figure 6:</u> SON accumulates at the mU2OS locus during the transition from transcriptionally active to inactive state.

Cells were transfected with pTetON and left uninduced (a-c), induced with dox for 5.5 h (d-f), or induced with dox for 2.5 h followed by media replacement wash without dox for 3 h (g-i). Cells were processed for immunolocalization of SON. Upper right value is PCC. Scale bar = 5 μm. Plot of pooled PCCs from n = 2 biological replicates and 30 cells scored for each condition per replicate. Bars denote mean and standard deviation. ** p-value = 0.0071; **** p-value < 0.0001 computed by one-way ANOVA and post-hoc Tukey's multiple comparisons procedure.


mCherry-LacI
SON
Merge
-dox
+dox
wash
a
b
c
d
e
f
g
h
i
0.332
0.005
0.322

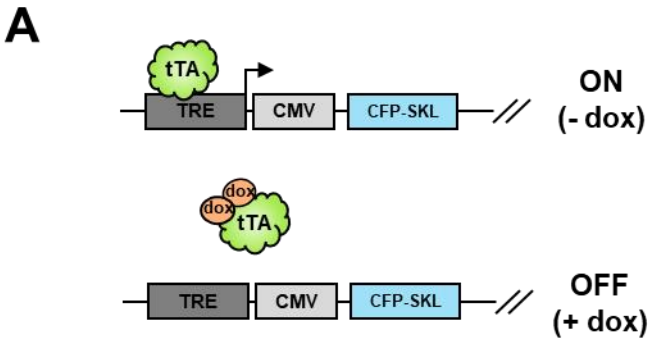

ns
**
****
1.0
0.5
0.0
-0.5
-1.0
PCC
Inactive
Active
Wash

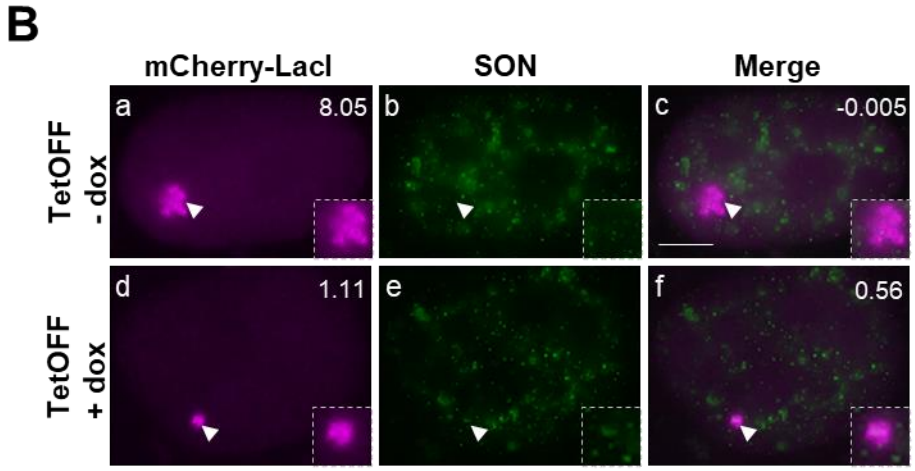
<u>Figure 7:</u> SON colocalizes with the TetOFF transcriptionally inactivated mU2OS locus.

A. Schematic representation of the TetOFF system. When pTetOFF is expressed in mU2OS in cells without doxycycline, the encoded tTA binds to the TREs and activates reporter gene transcription from the CMV promoter. In the presence of doxycycline, tTA cannot bind to the TREs, inactivating reporter gene expression.

B. SON did not colocalize with transcriptionally active loci transfected with TetOFF without dox (27/31 cells scored; panels a-c). SON colocalized with transcriptionally inactivated loci transfected with TetOFF in the presence of doxycycline (12/39 cells scored; panels d-f).

C. Heat map depicting the percentage of mU2OS loci with SON colocalization (PCC ≥ 0.1), no correlation ($-0.099 < PCC < 0.099$), or removed ($PCC \leq -0.1$) using the TetOFF system. Area of loci in pink are $< 2.0\ \mu m^2$, while loci in blue are $\geq 2.0\ \mu m^2$, bin = 10% of cells. Data represents PCCs from 31 cells (Mock, Active) or 53 cells (Inactivated) scored for n = 1 biological replicate.





C

	Mock	Active	Inactivated
Colocalized	54.839	12.903	22.642
No Correlation	19.355	3.226	26.415
Removed	22.581	9.677	24.528

Small: 51 - 60, 41 - 50, 31 - 40, 21 - 30, 11 - 20, 0 - 10

Colocalized	3.226	16.129	9.434
No Correlation	0	41.936	11.321
Removed	0	16.129	5.660

Large: 51 - 60, 41 - 50, 31 - 40, 21 - 30, 11 - 20, 0 - 10

Subaim 1b: Identify where and how SON interacts with the U2OS 2-6-3 locus.

SON associates with the promoter of the inactive mU2OS gene.

Microscopy analysis revealed SON localizes to the heterochromatic mU2OS locus, but which region of the reporter gene SON associates with and when is unknown. Chromatin immunoprecipitation (ChIP) with either anti-SON WU13 or no antibody control was performed on mU2OS lysate and the resulting purified DNA was analyzed by qPCR using primers recognizing the CMV promotor, 5' β-globin, or 3' β-globin regions of the U2OS 2-6-3 reporter gene array (schematic in Figure 8a). SON was preferentially enriched at the CMV promoter region in inactive cells (Figure 8b; pink). SON association with the CMV promoter was markedly reduced in mU2OS lysate from transcriptionally activated reporter cells (Figure 8b; blue). There was no change in enrichment of the 5' and 3' β-globin gene regions between transcription status, and reporter activated mU2OS lysate showed no change in enrichment at the reporter gene compared to no antibody control. This experiment suggests SON associates preferentially with the U2OS 2-6-3 promoter region and leaves the promoter upon reporter transcription activation.

SON's RNA associating domains are not required for mU2OS locus association.

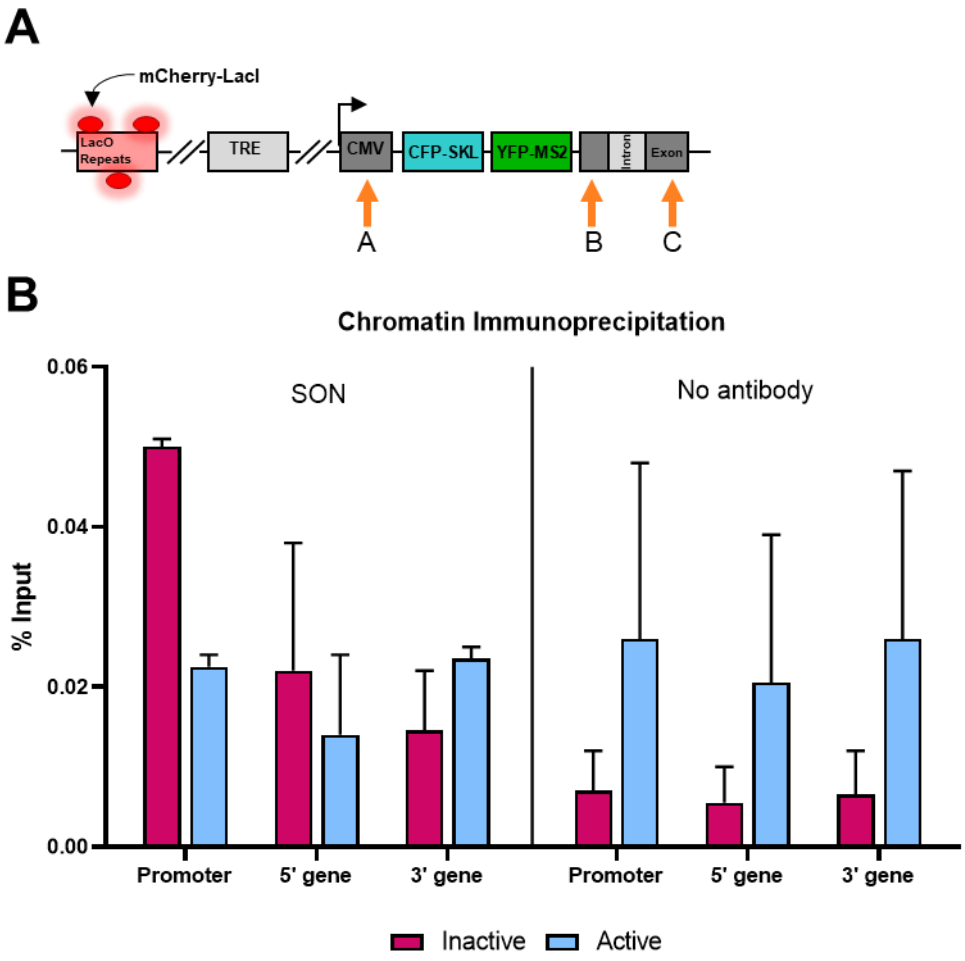
We were interested in which SON domains mediate its association with the inactive locus. The SON protein contains a large repetitive repeat motif domain, an RS domain involved in splicing, and a C-terminal G-patch and double-stranded RNA binding

motif (schematic in Figure 9a). siRNA refractory YFP-SON deletion mutants (Sharma et al., 2010) were transfected into U2OS 2-6-3 cells transiently expressing CFP-LacI. This experiment was optimized and executed by undergraduate researcher Jacob Ward. Jacob performed the experiment, I imaged the resulting slides, and Jacob scored the cells and created the figure (Figure 9b). Full-length YFP-SON (YFP-SON-FL) strongly colocalized with CFP-LacI in 100% of cells scored with an average PCC of 0.706 ± 0.230 (Figure 9b; panels a-c). SON containing the N-terminus through the unique tandem repeat motifs (YFP-SON 1-1493) also strongly colocalized with CFP-LacI in 80% of cells scored with an average PCC of 0.492 ± 0.457 (Figure 9b; panels f-h). Thus, deleting the RS domain, G-patch, and double stranded RNA binding motifs which are all known to facilitate pre-mRNA processing functions (Ahn et al., 2011), is not sufficient to alter SON colocalization with the heterochromatic reporter gene locus. This experiment also rules out anti-SON antibody cross-reactivity with the inactive reporter locus and demonstrates that exogenous YFP-SON also localizes at the heterochromatic U2OS 2-6-3 reporter locus.

<u>Figure 8:</u> SON associates with the promoter of inactive mU2OS cells.

A. Schematic of the mU2OS reporter gene array identifying the regions amplified by qPCR primers. Primer set A amplifies a region of the CMV minimal promoter, primer set B amplifies a region at the 5' beginning of the β-globin intron/exon module, and primer set C amplifies a region at the 3' end of the β-globin intron/exon module.

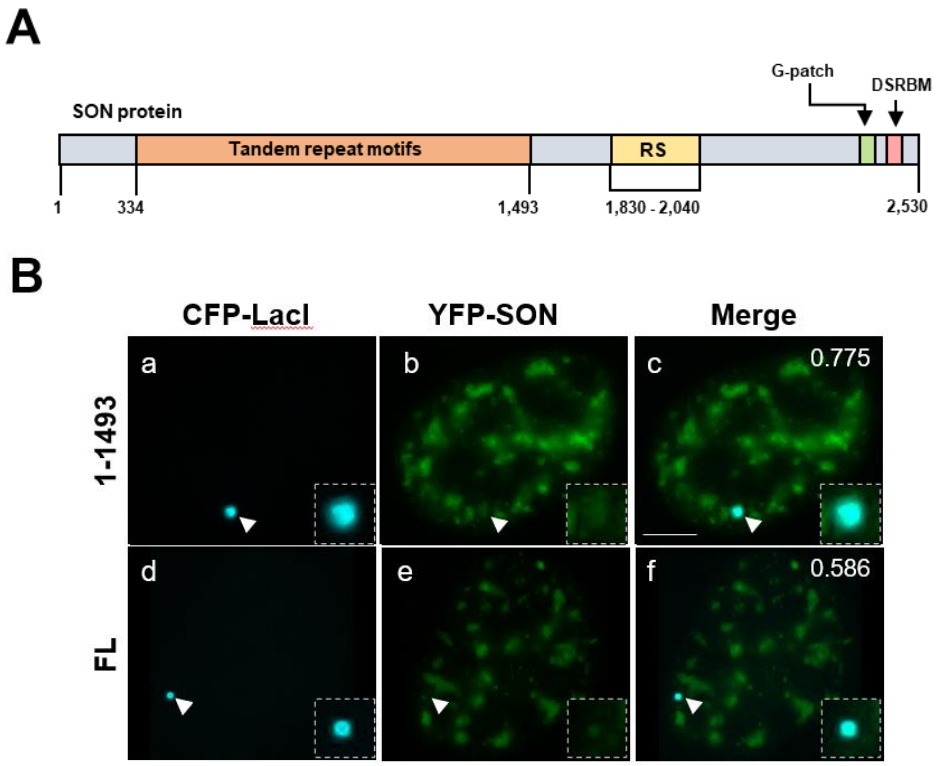
B. mU2OS cells transfected with carrier DNA only (inactive) or pTetON with doxycycline added (active) were harvested and SON pulled down by chromatin immunoprecipitation (ChIP). Anti-SON WU13 ChIP was compared to ChIP without antibody as a background signal control. One experimental unit consists of inactive and activated cells that were harvested at the same time and divided into anti-SON and no antibody control ChIP reactions. The experimental unit was repeated n = 2 times with 2 technical replicates per biological replicate. Error bars denote SEM.


A
mCherry-LacI
LacO Repeats
TRE
CMV
CFP-SKL
YFP-MS2
Intron
Exon
A
B
C
B
Chromatin Immunoprecipitation
SON
No antibody
% Input
0.06
0.04
0.02
0.00
Promoter
5' gene
3' gene
Promoter
5' gene
3' gene
Inactive
Active

<u>Figure 9:</u> SON repeat motifs are necessary for reporter locus colocalization

A. Schematic of SON protein domains.

B. Inactive reporter loci in U2OS 2-6-3 cells transfected with CFP-LacI and either YFP-SON 1-1493 (a-c; n = 25 cells scored) or YFP-SON FL (d-f; n = 26 cells scored). Pearson Correlation Coefficients of exogenous YFP-SON relative to CFP-LacI in upper right. Scale bar = 5 μm.


A
G-patch
DSRBM
SON protein
Tandem repeat motifs
RS
1
334
1,493
1,830 - 2,040
2,530
B
CFP-LacI
YFP-SON
Merge
1-1493
FL
a
b
c
0.775
d
e
f
0.586

CHAPTER V

AIM 2 RESULTS

Subaim 2a: Determine how SON affects chromatin organization of the U2OS 2-6-3 gene locus.

Our laboratory has demonstrated SON depletion results in the reorganization of nuclear speckle factors into a donut shape, showing a role for SON in nuclear speckle organization (Sharma et al., 2010). The repeat motifs are required for the donut-shaped phenotype and suggest a potential scaffold mechanism for SON in nuclear speckles. We were interested in defining the role of SON in chromatin organization. This subaim investigated SON's role in chromatin structure, and how SON depletion affects the chromatin organization of the U2OS 2-6-3 gene.

SON depletion affects chromatin organization of the mU2OS reporter gene.

mU2OS cells were treated with either non-targeting siRNA or SON siRNA duplexes to investigate how SON depletion affects the reporter locus. SON-depleted cells did not show the distinctive nuclear speckle staining pattern when immunolabeled with anti-SON WU13, indicating siRNA knock down of SON is detectable by microscopy (Figure 10a; panel e). In control cells, the reporter locus appears as a bright, condensed

dot indicative of the previously characterized heterochromatic state discussed in Aim 1 (Figure 10a; panels a-c). Strikingly, SON-depleted loci exhibited a significantly increased locus area compared to control loci (Figure 10b). The number of SON-depleted cells with decondensed loci (Figure 10a; panels d-f; 19/30 cells) was much higher than that in control siRNA-treated cells (Figure 10a; panels a-c; 1/30 cells). Decondensed SON-depleted loci appear morphologically rounded, similar to control inactive cell loci except with a larger area. This is in contrast to the irregularly shaped, globular, decondensed appearance that is a hallmark of transcriptionally active loci morphology.

It is important to reiterate that TetON was not expressed in control or SON-depleted conditions, so the increase in locus area cannot be due to reporter gene activation. The U2OS 2-6-3 reporter gene is tightly controlled by a tetracycline response element so that only expression of rtTA (rTetR fused to the VP16 activation domain; TetON) in the presence of doxycycline binds to the TRE to drive CMV minimal promoter activation and reporter gene transcription. Furthermore, immunolabeling of RNA pol II (Figure 11a) and pre-mRNA processing factor Btf (Figure 11b) showed no colocalization with the inactive reporter locus in control or SON-depleted cells. The absence of transcription factors at the inactive SON-depleted locus supports the conclusion that transcription is not occurring at these loci, despite the visible chromatin decondensation that is a hallmark of the transcriptionally activated U2OS 2-6-3 locus. Reporter transcript qRT-PCR analysis showed when SON mRNA is markedly reduced, reporter transcript is reduced compared to control cells (Figure 12). It is worth noting that normalization of reporter transcripts in SON-depleted cells compared to Control cells is a measure of relative quantification, not absolute quantification. Therefore, a reduction in reporter

transcript in SON-depleted cells compared to Control cells that already produce little to no reporter transcript is further indication that reporter transcript is not produced in SON-depleted cells.

SON-depleted loci are transcriptionally activatable.

Higher order chromatin decondensation is typically associated with transcription activation. Given that heterochromatic SON-depleted loci exhibit increased locus area without transcription activation, we next sought to determine if transcription activation could still occur normally in SON-depleted loci. mU2OS cells were treated with either non-targeting siRNA or SON siRNA duplexes before transfection with TetON and treatment with doxycycline to investigate how SON depletion affects the transcriptionally activated reporter locus. In control cells, the reporter locus appears large, decondensed, and with a globular morphology indicative of the previously characterized transcriptionally active state discussed in Aim 1 (Figure 13a; panels a-c). SON-depleted loci did not exhibit a change in locus area compared to control loci (Figure 13b). The number of SON-depleted cells with decondensed loci (Figure 13a; panels d-f; 22/31 cells) was about the same as in control siRNA-treated cells (Figure 13a; panels a-c; 23/30 cells). Figures 10 and 13 were performed in tandem. Morphologically, the decondensed SON-depleted locus varied, with some loci having a more rounded enlarged appearance, and some having a more irregular, globular appearance.

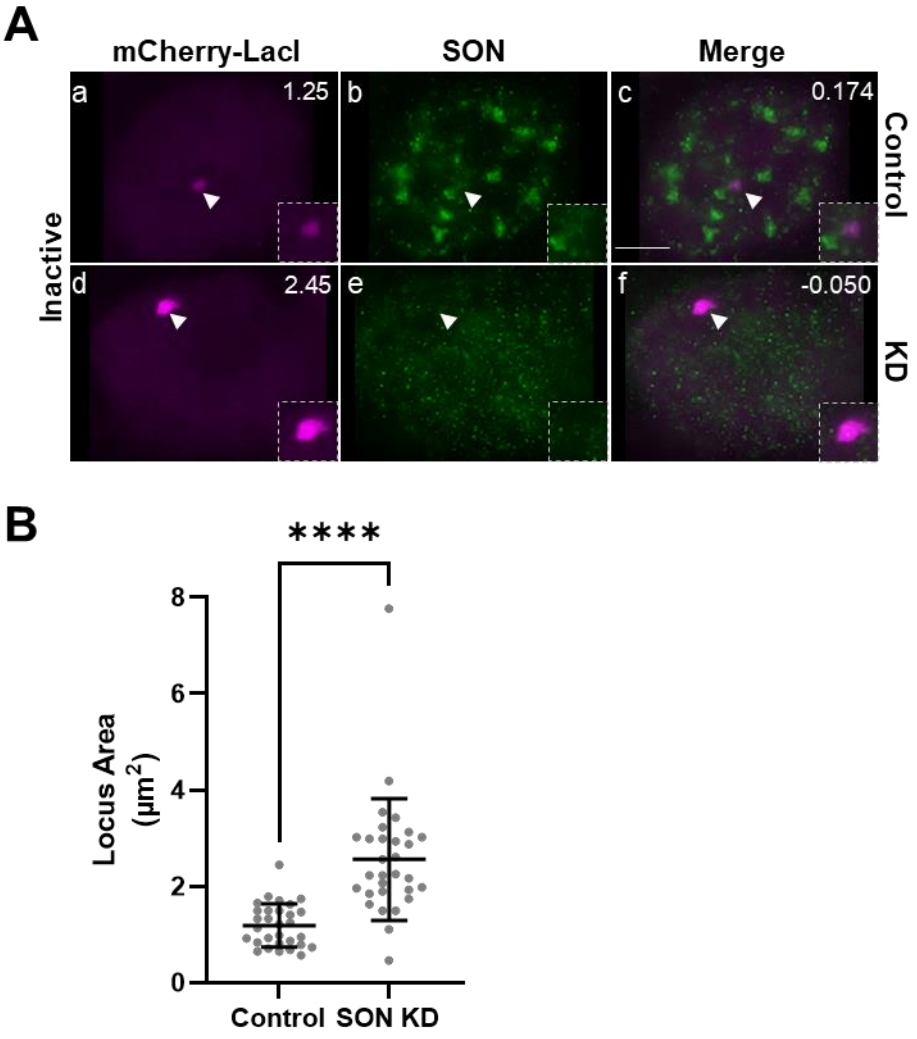
Transcription factors RNA pol II and Btf were immunolabelled and assessed for colocalization with the “TetON + dox” cells to determine if SON-depleted loci can be

transcriptionally active. Both RNA pol II (Figure 14a) and pre-mRNA processing factor Btf (Figure 14b) showed colocalization with the active reporter locus in control and SON-depleted cells. Figures 11 and 14 were performed in tandem. The presence of transcription factors at the transcriptionally active SON-depleted locus indicates transcription is still occurring at these loci. This is unsurprising since SON does not colocalize with control active cells (Aim 1) and previous work by our laboratory has shown that SON depletion does not alter global transcription levels (Sharma et al., 2011).

Figure 10. SON-depleted reporter loci exhibit increased locus area.

A. Reporter locus size was measured in mU2OS cells treated with either non-targeting siRNA duplexes (a-c) or SON siRNA duplexes 1 + 4 (d-f) and processed for immunolocalization of SON. Upper right value in LacI channel indicates locus area (μm^2); upper right value in merge channel indicates PCC. Scale bar = 5 μm.

B. Dot plot displaying area of reporter loci in control or SON KD condition. In each condition n = 30 cells scored from n = 1 biological replicate. Bars denote mean and standard deviation. **** P-value < 0.0001 calculated by two-way ANOVA with Sidak's multiple comparisons test.

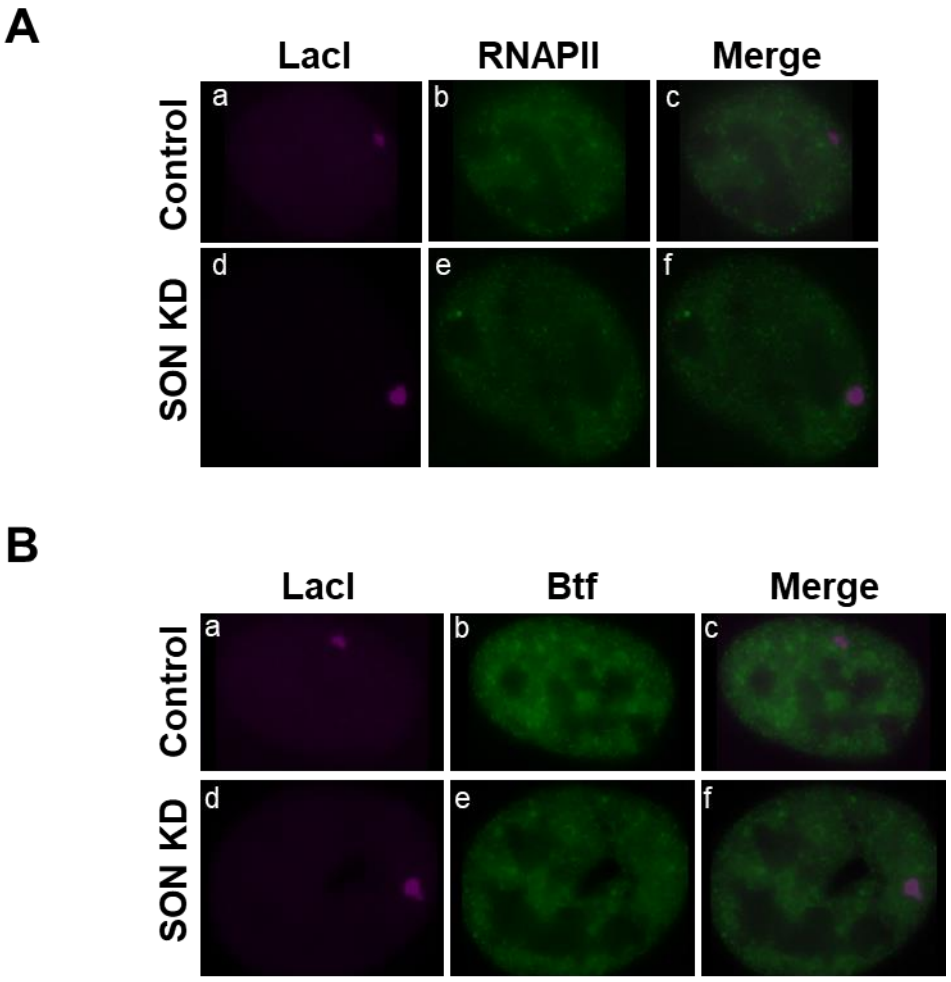

A
mCherry-LacI
SON
Merge
a
1.25
b
c
0.174
Control
Inactive
d
2.45
e
f
-0.050
KD
B
****
Locus Area (μm²)
8
6
4
2
0
Control
SON KD

<u>Figure 11:</u> Transcription and RNA processing factors do not colocalize with the SON-depleted, inactive decondensed reporter locus.

PCC was measured in mU2OS cells treated with either non-targeting siRNA duplexes (a-c) or SON siRNA duplexes 1 + 4 (d-f) and processed for immunolocalization of transcription factors. Upper right value in LacI channel indicates locus area (μm^2); upper right value in merge channel indicates PCC. Scale bar = 5 μm.

A. RNA pol-II does not colocalize with 26/30 control loci (panels a-c) and does not colocalize with 24/30 SON-depleted loci (panels d-f).

B. Pre-mRNA processing factor Btf does not colocalize with 30/33 control loci scored (panels a-c). Btf does not colocalize with 28/30 SON-depleted loci (panels d-f).


A
LacI
RNAPII
Merge
Control
SON KD
a
b
c
d
e
f
B
LacI
Btf
Merge
Control
SON KD
a
b
c
d
e
f

<u>Figure 12:</u> Reporter transcript is decreased when SON mRNA levels are reduced.

Relative mRNA expression of SON (left) or 2-6-3 transcript (right) from control or SON KD cells by qRT-PCR normalized to GAPDH and control cells. Data represents n = 3 biological replicates and error bars denote standard deviation.

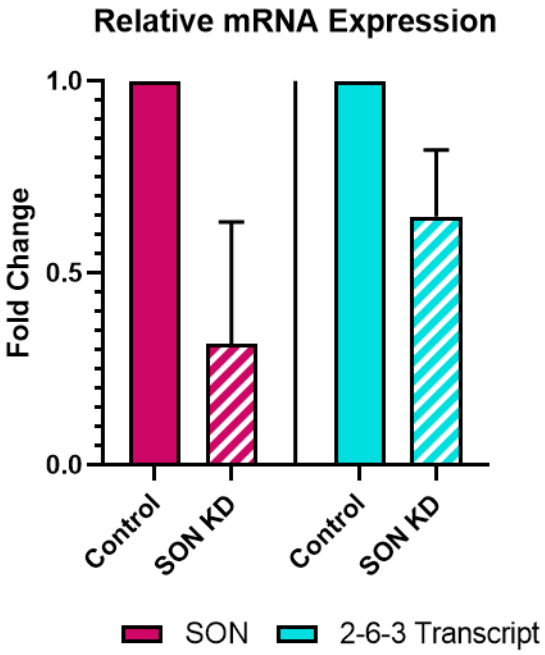
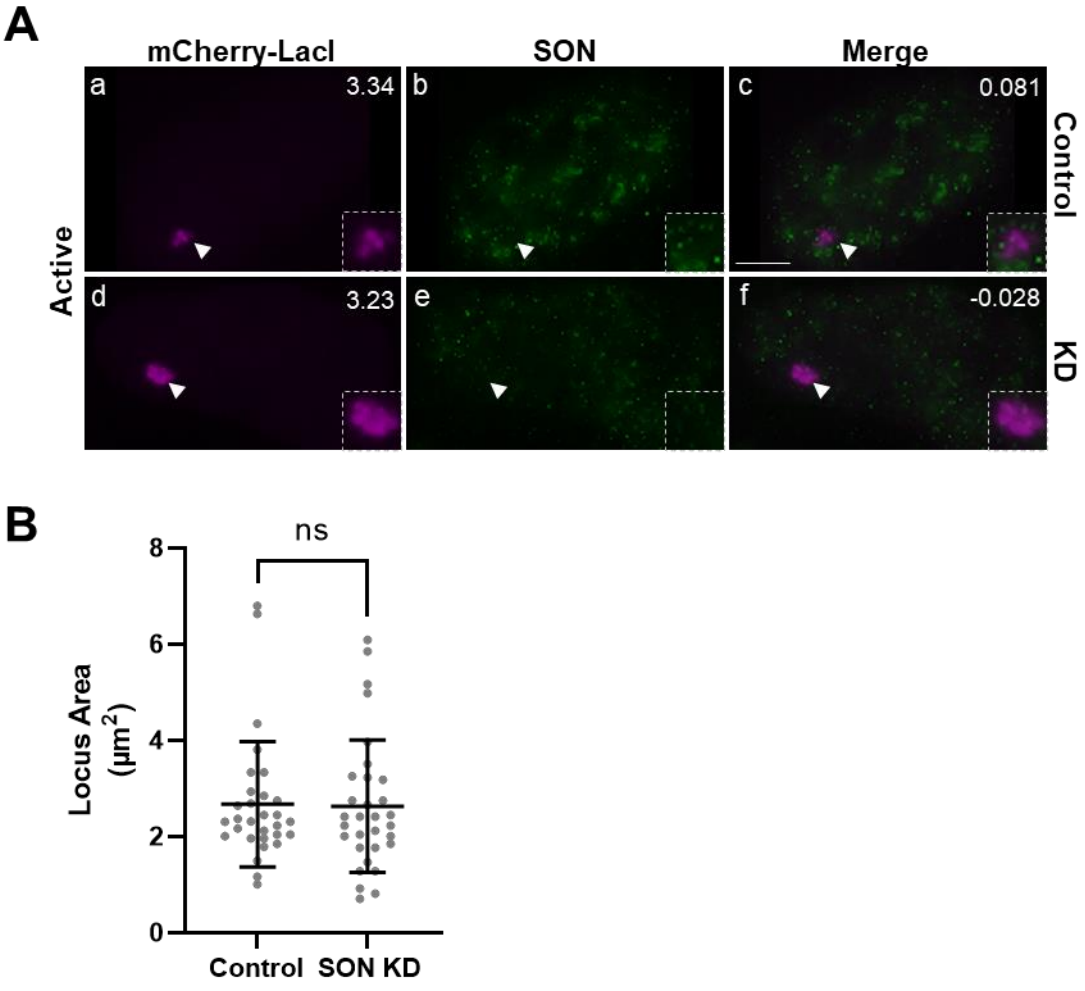

Relative mRNA Expression
Fold Change
1.0
0.5
0.0
Control
SON KD
Control
SON KD
SON
2-6-3 Transcript

Figure 13. SON depletion does not affect transcriptionally active locus area.

A. Reporter locus activation following treatment with non-targeting siRNA (a-c) or SON siRNA 1 + 4 (d-f) was done 48 hours after knockdown, as cells were electroporated with TetON, induced with dox +2.5 h, and processed for immunolocalization of SON. Upper right value in LacI channel indicates locus area (μm^2); upper right value in merge channel indicates PCC. Scale bar = 5 μm.

B. Dot plot displaying area of reporter loci in control (n = 30 cells) or SON KD condition (n = 31 cells). Bars denote mean and standard deviation. P-value > 0.9999 calculated by two-way ANOVA with Sidak's multiple comparisons test.

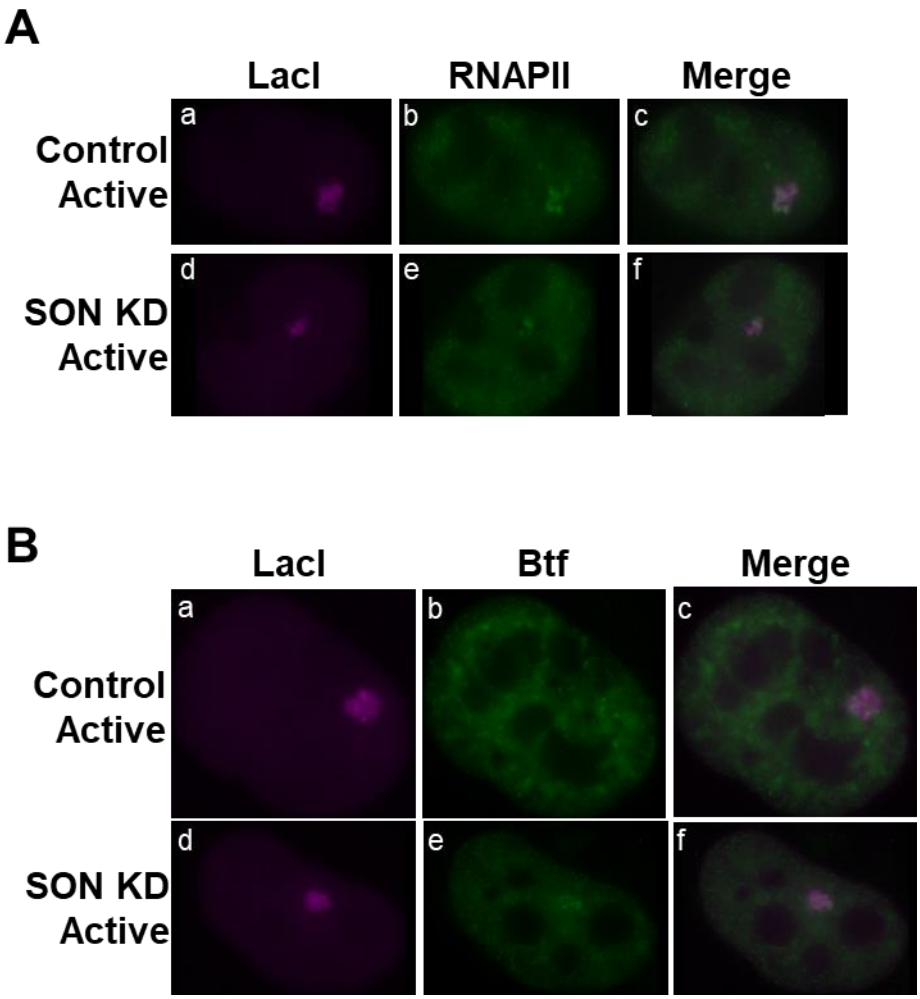

A
mCherry-LacI
SON
Merge
a
3.34
b
c
0.081
Control
Active
d
3.23
e
f
-0.028
KD
B
ns
Locus Area (μm²)
8
6
4
2
0
Control
SON KD

<u>Figure 14:</u> Transcription and RNA processing factors colocalize with the SON-depleted, active decondensed reporter locus.

PCC was measured in mU2OS cells treated with either non-targeting siRNA duplexes (a-c) or SON siRNA duplexes 1 + 4 (d-f) and processed for immunolocalization transcription factors. Upper right value in LacI channel indicates locus area (μm^2). Scale bar = 5 μm.

A. RNA pol-II colocalizes with 23/30 control loci (panels a-c) and colocalizes with 17/30 SON-depleted loci (panels d-f).

B. Pre-mRNA processing factor Btf colocalizes with 24/30 control loci scored (panels a-c) and colocalizes with 21/30 SON-depleted loci (panels d-f).


A
LacI
RNAPII
Merge
Control Active
SON KD Active
a
b
c
d
e
f
B
LacI
Btf
Merge
Control Active
SON KD Active
a
b
c
d
e
f

SON is required for reporter gene chromatin stability independent of transcription elongation.

Given that SON depletion alters reporter gene chromatin condensation without transcription activation, we next wanted to examine how global transcription inhibition affects the SON-depleted decondensed reporter locus. I used α-amanitin (aA), an irreversible transcription inhibitor that prevents RNA pol II nucleotide incorporation and leads to degradation of Rpb1 (Bensaude, 2011), to block global transcription elongation. When mU2OS cells are treated with α-amanitin, YFP-MS2BP is not recruited to the reporter locus because reporter mRNA is not being transcribed (Figure 15). If SON-depleted locus decondensation occurs separately from transcription elongation, inhibiting transcription via α-amanitin should not decrease the locus area. In control cells, we expected α-amanitin pre-treated cells to exhibit significantly smaller loci than transcriptionally active cells without α-amanitin (Rafalska-Metcalf et al., 2010; Raghuram et al., 2013; Tripathi et al., 2010; Wei et al., 2015). α-Amanitin-treated nuclei also display enlarged, rounded SON-labeled nuclear speckles due to inhibited pre-mRNA processing factor recruitment (Figure 16a; Bernard et al., 2010; Hu et al., 2010; Sharma et al., 2010). Control cells pre-treated with α-amanitin before transcription activation with dox exhibited a significantly lower mU2OS reporter locus area compared to untreated control cells (Figure 16b; left). There was no significant difference in locus area between SON-depleted cells with α-amanitin or without α-amanitin (Figure 16b; right). There was a significant decrease in locus area in SON-depleted α-amanitin cells compared to control active cells (Figure 16b; columns 1 and 4). This is not entirely unexpected, since pre-treatment with α-amanitin means loci never get transcriptionally activated, so it is

reasonable that SON-depleted loci would be somewhat smaller than fully decondensed transcriptionally active control loci without α-amanitin. Interestingly, SON-depleted α-amanitin loci were still significantly larger than control cell loci with α-amanitin (Figure 16b; columns 2 and 4). This supports our hypothesis that transcription inhibition alone is not sufficient to condense SON-depleted chromatin.

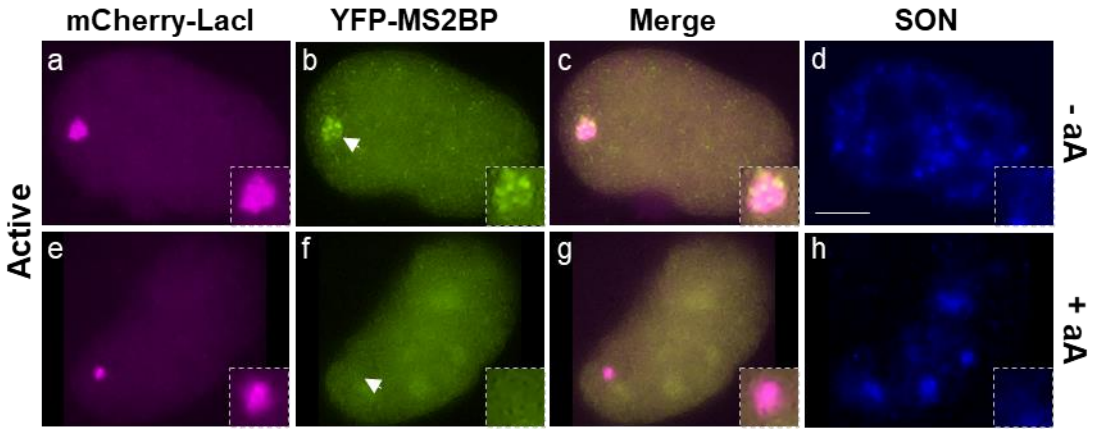
As a control to ensure that α-amanitin-induced reporter locus changes were due to transcription elongation inhibition, cells without TetON/dox were subjected to α-amanitin treatment and reporter loci area assessed (Figure 17). Control cells still exhibited the enlarged speckle staining pattern indicative of α-amanitin transcription inhibition (Figure 17a; panels a-c) and did not display a significant decrease in reporter locus area following α-amanitin treatment (Figure 17b; left). Since untreated control cells are already transcriptionally inactive and highly condensed, we would not expect a further condensation in inactive α-amanitin treated control cells. When SON is depleted (Figure 17a; panels d-f), α-amanitin treated cells also did not show a significant difference in reporter locus area compared to untreated SON-depleted loci (Figure 17b; right). Since α-amanitin is added to the cells after SON depletion, and the SON-depleted locus remains transcriptionally inactive yet decondensed, it is unsurprising that loci are decondensed to the same extent in both treated and untreated SON-depleted conditions.

Experimental setup for Figures 16 and 17 were performed in tandem as indicated in Materials and Methods and repeated for n = 2 biological replicates. The compiled locus area scoring data with complete pairwise comparison analysis is displayed in Figure 18 and p-values tabulated in Table 2. The compiled data gives further insight into SON's role in chromatin decondensation in relation to transcription status. SON-depleted

uninduced loci (with or without α-amanitin treatment) are significantly smaller than control active loci. This suggests the locus decondensation observed in SON-depleted cells is not as extensive as the large-scale chromatin opening associated with transcription activation (Figure 18; column 3 vs. 5 and column 4 vs. 5). Furthermore, SON-depleted uninduced loci are significantly smaller than loci in SON-depleted active cells (Figure 18; column 3 vs. 7 and column 4 vs. 7), which supports the observation that SON-depleted loci are still transcriptionally activatable (Figures 13 and 14), resulting in further chromatin decondensation. Finally, SON-depleted "Active" loci pre-treated with α-amanitin are still significantly larger than loci in control inactive cells, indicating SON depletion, regardless of transcription status, is sufficient for significant increases in chromatin decondensation (Figure 18; column 1 vs. 8 and column 2 vs. 8). Taken together, these results indicate that transcription elongation inhibition by α-amanitin is not sufficient to condense the enlarged SON-depleted mU2OS locus, supporting the view that SON-depleted loci are decondensed via a mechanism that is uncoupled from transcription.

Figure 15: α-Amanitin inhibits reporter mRNA synthesis.

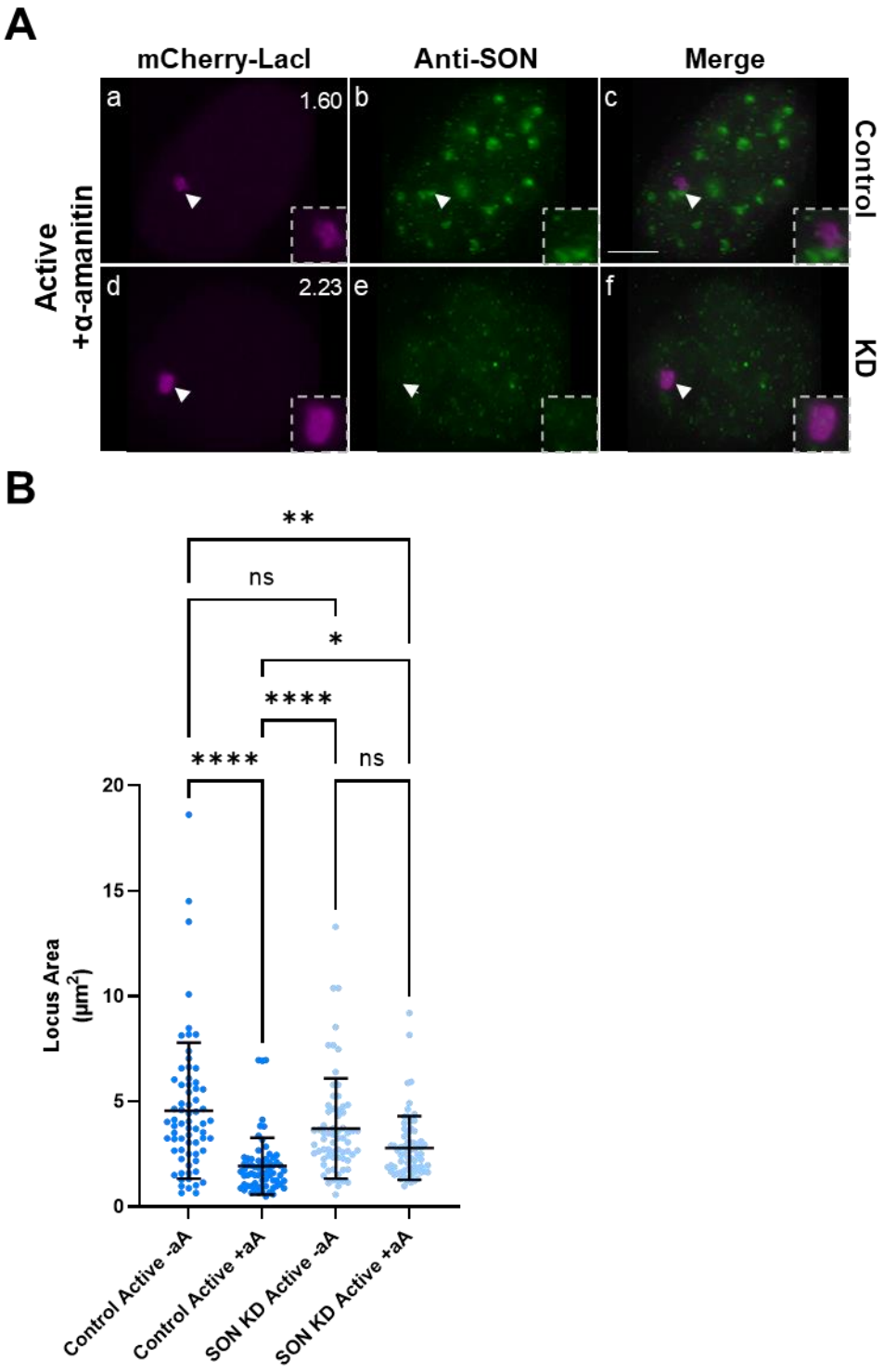
mU2OS cells were transfected with TetON and YFP-MS2BP to label reporter mRNA. Post-transfection 3 hrs, cells were either left untreated (panels a-d) or pre-treated with α-amanitin (panels e-h), then dox added to both conditions 3 hrs post-amanitin treatment. A total of 9 hrs post-transfection, cells were processed for immunofluorescence with anti-SON WU13 primary and DyLight-405 secondary antibody. Untreated active cells show YFP-MS2BP enrichment at the decondensed locus and normal SON-labeled nuclear speckle morphology, while α-amanitin treated cells do not recruit YFP-MS2BP and display enlarged SON-labeled nuclear speckles. Scale bar = 5 μm.


mCherry-LacI
YFP-MS2BP
Merge
SON
Active
a
b
c
d
e
f
g
h
- aA
+ aA

<u>Figure 16:</u> Transcription elongation inhibition does not alter SON-depleted locus area.

A. mU2OS cells with non-targeting siRNA (panels a-c) or SON siRNA 1 + 4 (panels d-f) were transfected with TetON and treated with 6 hrs α-amanitin (3 hrs pre-dox + 3 hrs post-dox) to inhibit global transcription elongation. Upper right value in LacI channel indicates locus area (μm^2). Scale bar = 5 μm.

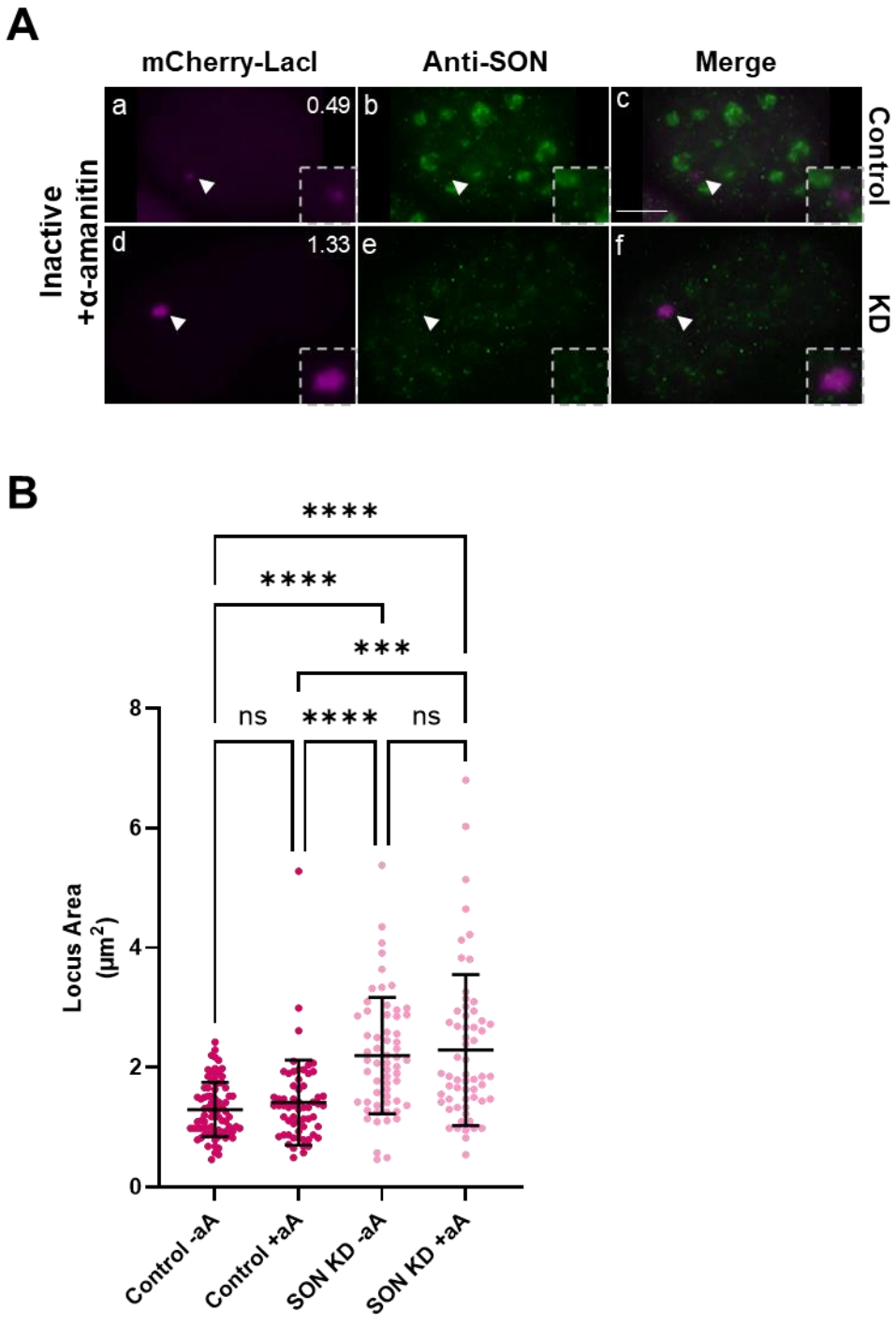
B. Dot plot displaying area of reporter loci in control or SON KD condition. Data represents 30 cells scored per condition and pooled from n = 2 biological replicates. Bars denote mean and standard deviation. * p-value < 0.05; ** p-value = 0.005; **** p-value < 0.0001 calculated by three-way ANOVA with Games Howell procedure, followed by Tukey's multiple comparisons test.


A
mCherry-LacI
Anti-SON
Merge
a
1.60
b
c
Control
Active
+α-amanitin
d
2.23
e
f
KD
B
**
ns
*
****
****
ns
20
15
10
5
0
Locus Area
(μm²)
Control Active -aA
Control Active +aA
SON KD Active -aA
SON KD Active +aA

<u>Figure 17:</u> Transcription elongation inhibition does not alter locus area in uninduced mU2OS cells.

A. mU2OS cells with non-targeting siRNA (panels a-c) or SON siRNA 1 + 4 (panels d-f) were left uninduced and treated with 6 hrs α-amanitin to inhibit global transcription elongation. Upper right value in LacI channel indicates locus area (μm^2). Scale bar = 5 μm.

B. Dot plot displaying area of reporter loci in control untreated (n = 80 cells), control aA-treated (n = 61 cells), SON KD untreated (n = 60 cells), or SON KD aA-treated (n = 60 cells) condition. Data was pooled from n = 2 biological replicates. Bars denote mean and standard deviation. *** p-value < 0.0005; **** p-value < 0.0001 calculated by three-way ANOVA with Games Howell procedure, followed by Tukey's multiple comparisons test.


A
mCherry-LacI
Anti-SON
Merge
Inactive
+α-amanitin
Control
KD
a
0.49
b
c
d
1.33
e
f
B
****
****
***
ns
****
ns
Locus Area (µm²)
8
6
4
2
0
Control -aA
Control +aA
SON KD -aA
SON KD +aA

<u>Figure 18:</u> Compiled locus area scoring data with pairwise comparisons from Figures 16 and 17.

These data are identical to the data displayed in Figures 16 and 17, compiled because all 8 cell conditions shown were performed together for n = 2 biological replicates. Violin plots displaying distribution data; bars indicate mean and upper/lower quartiles.

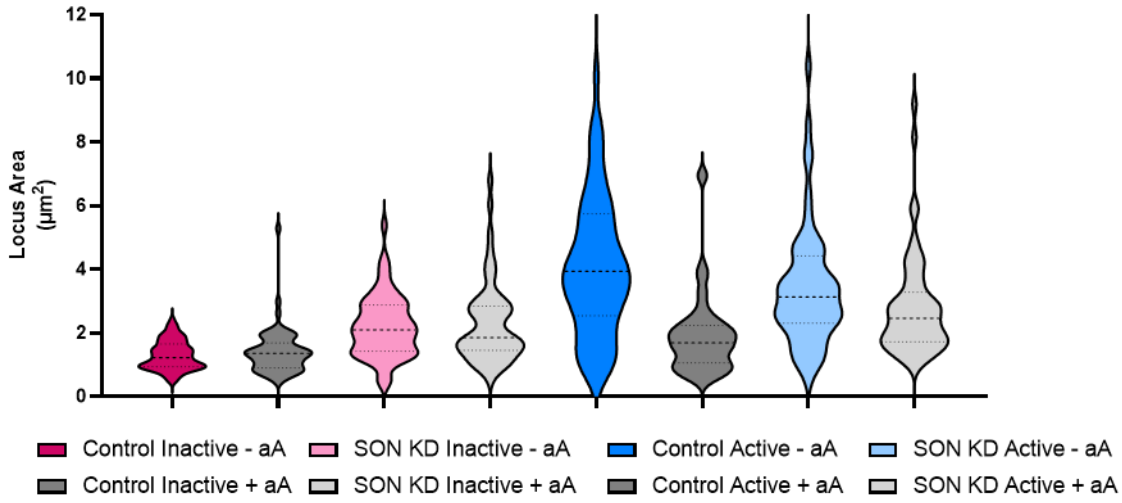

Locus Area (µm²)
12
10
8
6
4
2
0
Control Inactive - aA
SON KD Inactive - aA
Control Active - aA
SON KD Active - aA
Control Inactive + aA
SON KD Inactive + aA
Control Active + aA
SON KD Active + aA

Table 2: Tukey's multiple comparisons test of data from Figures 16-18.

Statistical analysis for Figures 16-18 was performed together, computed from three-way ANOVA with Games Howell procedure followed by Tukey's multiple comparisons test.

Comparison	Difference	p-value	95% CI
Control Active No - Control Active Yes	**2.70**	**< .0001**	**(1.3, 4.1)**
Control Active No - Control Inactive No	**3.31**	**< .0001**	**(1.99, 4.63)**
Control Active No - Control Inactive Yes	**3.19**	**< .0001**	**(1.85, 4.54)**
Control Active No - Depleted Active No	1.03	0.49	(-0.55, 2.62)
Control Active No - Depleted Active Yes	**1.80**	**0.0046**	**(0.35, 3.26)**
Control Active No - Depleted Inactive No	**2.41**	**< .0001**	**(1.04, 3.77)**
Control Active No - Depleted Inactive Yes	**2.31**	**< .0001**	**(0.91, 3.72)**
Control Active Yes - Control Inactive No	**0.61**	**0.0092**	**(0.09, 1.13)**
Control Active Yes - Control Inactive Yes	0.50	0.14	(-0.07, 1.07)
Control Active Yes - Depleted Active No	**-1.66**	**< .0001**	**(-2.68, -0.64)**
Control Active Yes - Depleted Active Yes	**-0.89**	**0.0165**	**(-1.69, -0.1)**
Control Active Yes - Depleted Inactive No	-0.29	0.85	(-0.92, 0.34)
Control Active Yes - Depleted Inactive Yes	-0.38	0.71	(-1.09, 0.32)
Control Inactive No - Control Inactive Yes	-0.11	0.96	(-0.43, 0.21)
Control Inactive No - Depleted Active No	**-2.27**	**< .0001**	**(-3.18, -1.37)**
Control Inactive No - Depleted Active Yes	**-1.50**	**< .0001**	**(-2.15, -0.86)**
Control Inactive No - Depleted Inactive No	**-0.90**	**< .0001**	**(-1.32, -0.49)**
Control Inactive No - Depleted Inactive Yes	**-0.99**	**< .0001**	**(-1.52, -0.47)**
Control Inactive Yes - Depleted Active No	**-2.16**	**< .0001**	**(-3.1, -1.22)**
Control Inactive Yes - Depleted Active Yes	**-1.39**	**< .0001**	**(-2.07, -0.71)**
Control Inactive Yes - Depleted Inactive No	**-0.79**	**< .0001**	**(-1.26, -0.31)**
Control Inactive Yes - Depleted Inactive Yes	**-0.88**	**0.0001**	**(-1.45, -0.31)**
Depleted Active No - Depleted Active Yes	0.77	0.38	(-0.32, 1.86)
Depleted Active No - Depleted Inactive No	**1.37**	**0.0006**	**(0.4, 2.35)**
Depleted Active No - Depleted Inactive Yes	**1.28**	**0.0042**	**(0.26, 2.3)**
Depleted Active Yes - Depleted Inactive No	0.60	0.19	(-0.13, 1.34)
Depleted Active Yes - Depleted Inactive Yes	0.51	0.52	(-0.29, 1.31)
Depleted Inactive No - Depleted Inactive Yes	-0.09	0.99	(-0.72, 0.54)

Subaim 2b: Determine how SON depletion affects the chromatin landscape and localization of chromatin modifiers at the mU2OS locus.

A ChIP-seq analysis by Kim et al. (2016) identified that SON associates with gene promoters and SON depletion is correlated with altered chromatin markers at these gene promoters. DNA is generally tightly wound around histones and packaged into highly condensed fibers with limited accessibility to transcription factors (Allshire et al., 2018; Eskeland et al., 2010; Ishihara et al., 2021). Chromatin decondensation via histone acetylation, DNA methylation, and/or chromatin remodeling factors increases chromatin accessibility and allows for transcription factor binding and transcription initiation. An increase in chromatin accessibility is correlated with an increase in gene expression (Gorisch et al., 2005; Esmaeili et al., 2020). Given SON's position at gene promoters, a highly accessible chromatin region, we were interested in SON's role in global chromatin structure and accessibility.

The condensed, transcriptionally silent reporter locus has been shown to behave like heterochromatin, exhibiting HP1 association as well as heterochromatic epigenetic marker H3K9me3 and its histone methyl transferases Suv39h1 and G9a-L (Janicki et al., 2004). Upon transcription activation with TetON + dox, these heterochromatic markers are depleted from the locus, corresponding with higher order chromatin decondensation and the deposition of histone variant H3.3 (Janicki et al., 2004). Previous reports have suggested that SON is a transcriptional regulator at gene promoters (Karlas et al., 2010; Kim et al., 2016), and that SON associates with the inactive U2OS 2-6-3 locus (Sharma 2011, pg. 147; this document, Aim 1), so we wanted to investigate the epigenetic landscape of SON-associated chromatin.

SON depletion alters epigenetic marker H3K9me3 localization at the reporter locus.

I first replicated the findings from Janicki et al. (2004) that H3K9me3 primarily colocalizes with the inactive, but largely not the active, reporter locus in mU2OS cells (Figure 19a; scoring data displayed in Figure 20). For epigenetic protein colocalization, a higher threshold PCC $\geq$ 0.20 was used to designate colocalization due to the diffuse nuclear localization of these proteins resulting in inflated colocalization values. Chromatin immunoprecipitation using primers designed along the reporter gene also showed enrichment of H3K9me3 along the WT mU2OS reporter gene (Figure 19b).

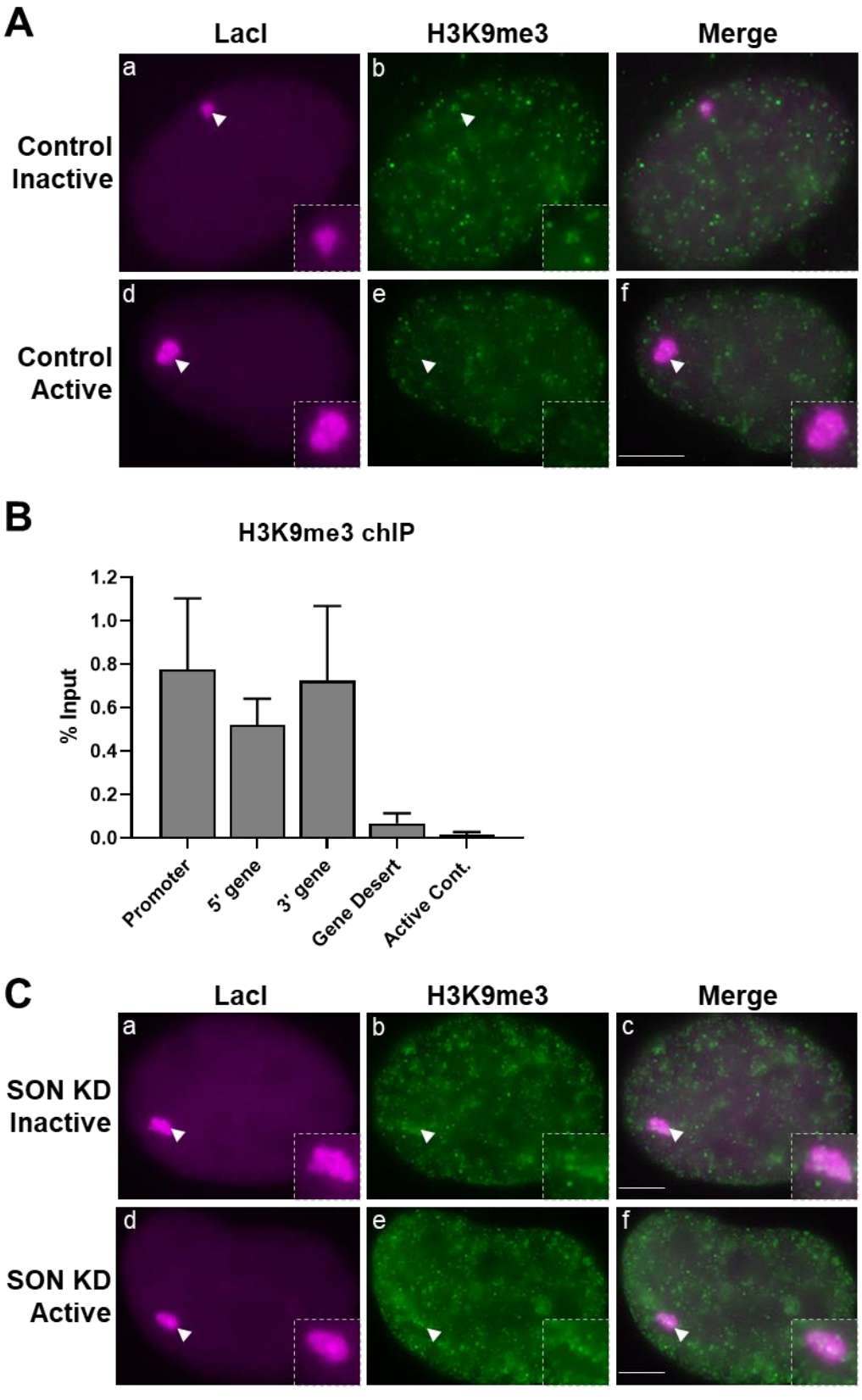
Interestingly, SON-depleted loci exhibited a slightly smaller proportion of cells with H3K9me3 colocalization in the inactive condition, and strong colocalization with the transcriptionally active, SON-depleted loci (images in Figure 19c; scoring data for images in Figure 19 displayed in Figure 20). The proportion of control active and SON-depleted inactive loci showing colocalization or no correlation of H3K9me3 was all roughly the same so perhaps completing more biological replicates or scoring more cells per condition would be necessary for more conclusive H3K9me3 localization status in these conditions. In addition to an increase in the number of scored cells exhibiting H3K9me3 colocalization with mCherry-LacI in SON-depleted active cells, pairwise comparison of H3K9me3 colocalization with mCherry-LacI showed a significant increase in the extent of H3K9me3 colocalization (significantly higher average PCC) at SON-depleted transcriptionally active loci compared to all other conditions (Figure 20b). The strong colocalization of the heterochromatic H3K9me3 marker with the euchromatic SON-depleted locus is striking and may suggest that SON maintains chromatin stability via H3K9me3 histone modification regulation.

<u>Figure 19:</u> H3K9me3 localization is altered following SON depletion.

A. Colocalization with mCherry-LacI was measured in mU2OS cells treated with non-targeting siRNA duplexes and left uninduced (panels a-c) or induced with TetON + dox (panels d-f) and processed for immunolocalization of H3K9me3. Scale bar = 5 μm.

B. WT mU2OS cells were harvested and H3K9me3 pulled down by chromatin immunoprecipitation (ChIP). Primers amplifying a region of the U2OS 2-6-3 CMV promoter, the 5' beginning of the β-globin intron/exon module, and a region at the 3' end of the β-globin intron/exon module were run alongside genomic controls amplifying a region of the OR4F5 gene desert and RNA Pol II active promoter region. The experiment was repeated n = 5 times with 2 technical replicates per biological replicate. Error bars denote SEM.

C. Colocalization with mCherry-LacI was measured in mU2OS cells treated with SON siRNA 1 + 4 duplexes and left uninduced (panels a-c) or induced with TetON + dox (panels d-f) and processed for immunolocalization of H3K9me3. Scale bar = 5 μm.

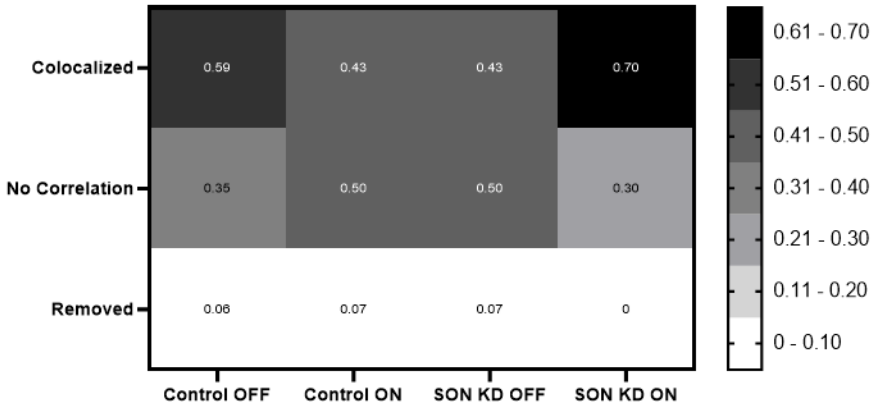

A
LacI
H3K9me3
Merge
Control Inactive
Control Active
a
b
d
e
f
B
H3K9me3 chIP
% Input
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Promoter
5' gene
3' gene
Gene Desert
Active Cont.
C
LacI
H3K9me3
Merge
SON KD Inactive
SON KD Active
a
b
c
d
e
f

<u>Figure 20:</u> There is an increase in H3K9me3 recruitment to the transcriptionally active SON-depleted reporter gene locus.

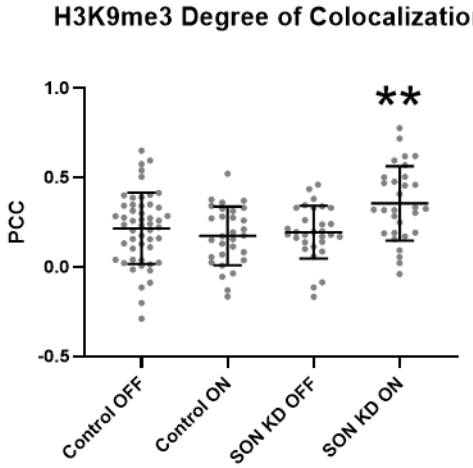
A. Heat map indicating proportion of loci scored per condition with H3K9me3 colocalized with mCherry LacI (PCC ≥ 0.2), no correlation (-0.1 < PCC <0.2) or removed (PCC ≤ -0.1) relative to locus condensation state. Bin = 0.1 (10%) cells. Data represents PCCs of 51 control inactive cells and 30 cells each scored in the remaining conditions.

B. Dot plot displaying PCCs of H3K9me3 vs. mCherry-LacI colocalization from the scored data in Figure 20a above. ** P-value < 0.005 calculated by two-way ANOVA followed by Tukey's multiple comparisons test.

A


H3K9me3 Colocalization with mCherry-LacI
Colocalized
No Correlation
Removed
0.59
0.43
0.43
0.70
0.35
0.50
0.50
0.30
0.06
0.07
0.07
0
Control OFF
Control ON
SON KD OFF
SON KD ON
0.61 - 0.70
0.51 - 0.60
0.41 - 0.50
0.31 - 0.40
0.21 - 0.30
0.11 - 0.20
0 - 0.10

B


H3K9me3 Degree of Colocalization
PCC
1.0
0.5
0.0
-0.5
**
Control OFF
Control ON
SON KD OFF
SON KD ON

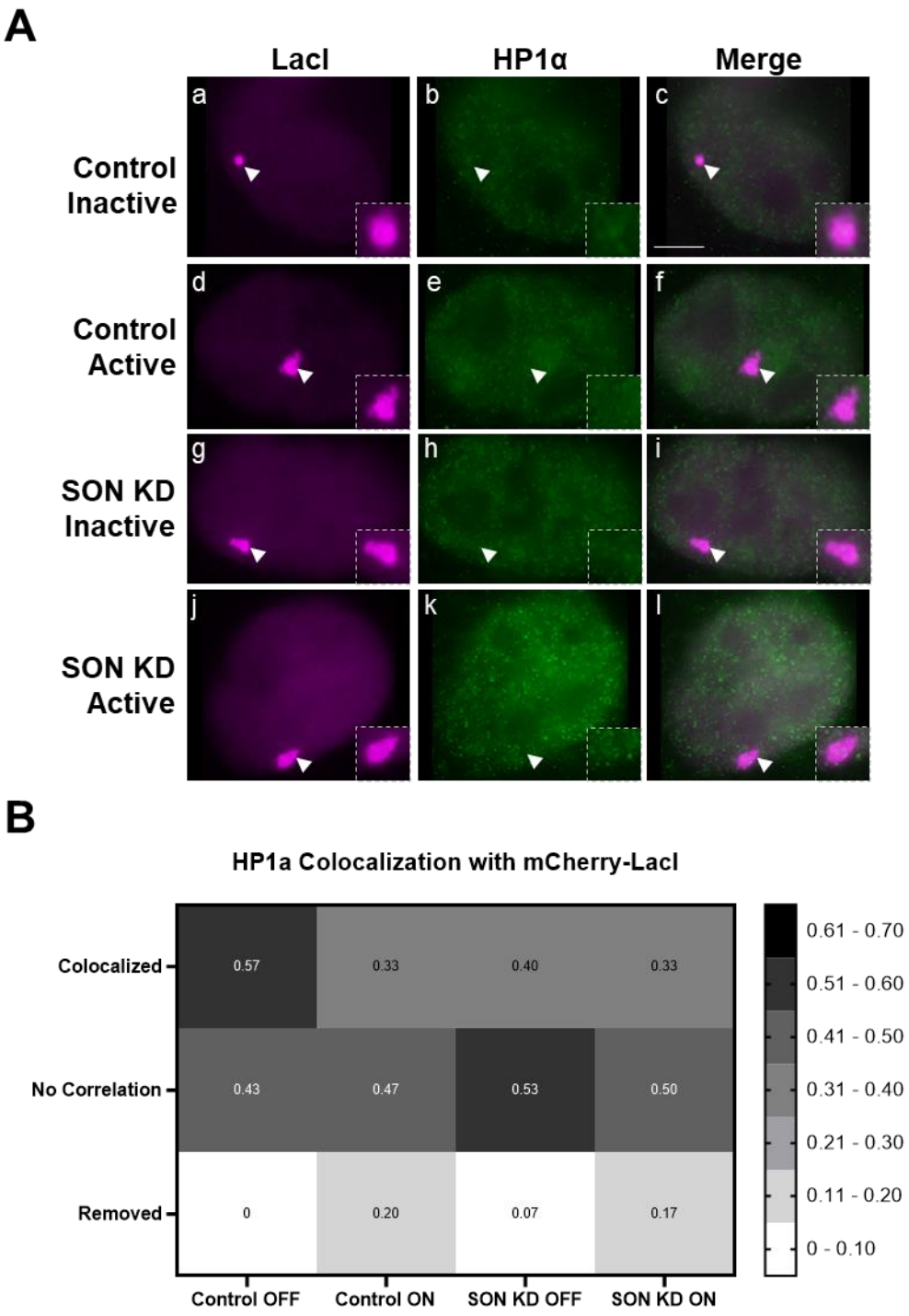
HP1α is a critical protein in the stability of heterochromatin and binds to H3K9me3, so we were interested if HP1α localization was also altered at the SON-depleted reporter locus. In control cells, I corroborated that HP1α colocalizes with the heterochromatic locus and is depleted upon transcription activation (Janicki et al., 2004; Figure 21a panels a-f; scoring in Figure 21b). However, HP1α colocalization with the SON-depleted inactive locus was inconclusive, with a similar number of cells exhibiting HP1α colocalization and no correlation with mCherry-LacI (images in Figure 21a; panels g-l; scoring in Figure 21b). HP1α was not colocalized with the SON-depleted active locus, although this is to be expected since HP1α is depleted from transcriptionally active loci (Janicki et al., 2004). More biological replicates or more cells scored per condition may be necessary to fully understand HP1α dynamics at the reporter locus, or colocalization analysis using exogenous YFP-HP1α expression rather than commercially available antibodies.

SON depletion is correlated with an increase in H3K4me3 at some promoter regions (Kim et al., 2016), so H3K4me3 localization at the SON-depleted reporter locus was analyzed. At this particular gene, H3K4me3 is not enriched at the inactive mU2OS locus, as evidenced by an average correlation of ~0 with mCherry-LacI via microscopy (Figure 22a). Likewise, SON depletion did not result in a change in colocalization of H3K4me3 at the mU2OS locus (Figure 22a). H3K4me3 was also not enriched at the reporter locus via chromatin immunoprecipitation analysis, and instead associated with the constitutively active ACTB-2 promoter region (Zhao et al., 2011; Figure 22b).

<u>Figure 21:</u> HP1α localization is not significantly altered following SON depletion.

A. Colocalization with mCherry-LacI was measured in mU2OS cells treated with non-targeting siRNA duplexes or SON si1 + si4 duplexes and left uninduced (panels a-c and panels g-i) or induced with TetON + dox (panels d-f and j-l) and processed for immunolocalization of HP1α. Scale bar = 5 μm.

B. Heat map indicating proportion of loci scored per condition with HP1α colocalized with mCherry LacI (PCC ≥ 0.2), no correlation (-0.1 < PCC <0.2) or removed (PCC ≤ -0.1) relative to locus condensation state. Bin = 0.1 (10%) cells. Data represents PCCs of 30 cells scored per condition.

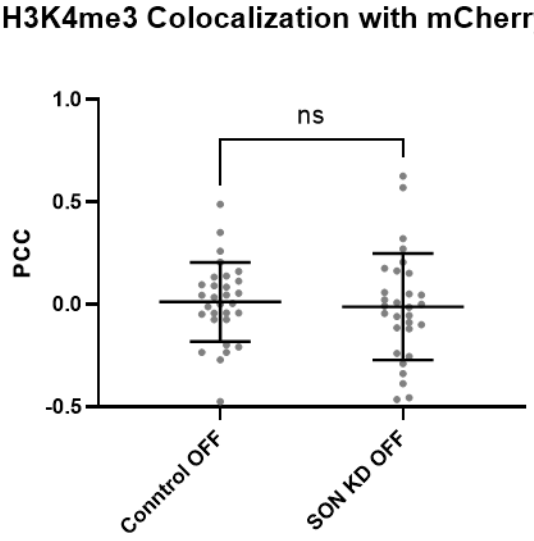

A
LacI
HP1α
Merge
Control Inactive
Control Active
SON KD Inactive
SON KD Active
a
b
c
d
e
f
g
h
i
j
k
l
B
HP1a Colocalization with mCherry-LacI
Colocalized
No Correlation
Removed
0.57
0.33
0.40
0.33
0.43
0.47
0.53
0.50
0
0.20
0.07
0.17
Control OFF
Control ON
SON KD OFF
SON KD ON
0.61 - 0.70
0.51 - 0.60
0.41 - 0.50
0.31 - 0.40
0.21 - 0.30
0.11 - 0.20
0 - 0.10

<u>Figure 22:</u> SON depletion does not alter H3K4me3 localization at the mU2OS locus.

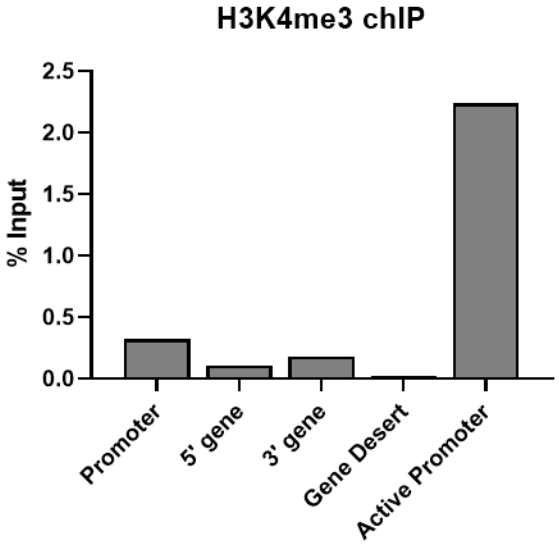
A. Dot plot displaying PCCs of H3K4me3 vs. mCherry-LacI colocalization from mU2OS cells treated with non-targeting siRNA duplexes (left) or SON si1 + si4 duplexes (right) without reporter gene transcription induction. P-value > 0.05 calculated by Welch's unpaired T-test.

B. WT mU2OS cells were harvested and H3K4me3 pulled down by chromatin immunoprecipitation (ChIP). Primers amplifying a region of the U2OS 2-6-3 CMV promoter, the 5' beginning of the β-globin intron/exon module, and a region at the 3' end of the β-globin intron/exon module were run alongside genomic controls amplifying a region of the OR4F5 gene desert and ACTB-2 active promoter region. The data represents n = 1 biological replicate with 2 technical replicates per biological replicate.

A


H3K4me3 Colocalization with mCherry-LacI
ns
PCC
1.0
0.5
0.0
-0.5
Conntrol OFF
SON KD OFF

B


H3K4me3 chIP
% Input
2.5
2.0
1.5
1.0
0.5
0.0
Promoter
5' gene
3' gene
Gene Desert
Active Promoter

Subaim 2c: Determine SON's role in genome-wide chromatin accessibility

In the previous Aim 2 subaims, I have uncovered a role for SON in maintaining chromatin stability at the mU2OS locus, observed that SON-dependent chromatin decondensation cannot be overcome by global transcription inhibition via α-amanitin treatment, and identified chromatin markers that exhibit altered localization following SON depletion. Given SON's role in chromatin condensation and its association with gene promoters, I was interested if SON could be involved in chromatin accessibility. Large-scale changes in chromatin accessibility can alter the 3D organization of chromatin which in turn can alter the long-range chromatin looping interactions of gene expression machinery. Genome-wide chromatin organization and its effects on the cell and human disease are currently largely unknown and the subject of an entire 4D Nucleome Consortium whose research is rapidly evolving (4D Nucleome, 2021). In this subaim, I investigated SON's role in chromatin accessibility via gene-specific DNase-qPCR and genome-wide DAPI analysis.

SON-depleted cells are more susceptible to DNase digestion.

Gene promoters are highly accessible regions of chromatin, meaning they are more decondensed to allow transcription factor binding, and physically have more distance between histones than heterochromatic regions. This open chromatin conformation is detectable by subjecting nuclei to mild DNase I digestion and analyzing the remaining purified DNA by agarose gel electrophoresis and gene locus-specific qPCR. More protected regions of chromatin are more resistant to DNase digestion and

interpreted as being more condensed, whereas less protected chromatin is more susceptible to DNase digestion and is interpreted as a more accessible region. I first conducted a pilot study to determine the amount of DNase I enzyme necessary to partially digest chromatin from mU2OS cells without completely degrading the DNA sample. Nuclei from mU2OS cells were isolated and subjected to varying DNase I reaction temperatures and times. Following mild chromatin sonication and DNA purification, samples were analyzed by agarose gel electrophoresis and 65° C for 5 min was determined as the best DNase I condition for marked reduction in DNA quantity and overall shorter DNA fragments compared to uncut (DNase I-free) condition (Figure 23a).

Analysis of chromatin accessibility changes at specific gene loci can be assessed quantitatively based on the principle that more accessible regions are more degraded by DNase I which yields less input substrate, thereby require more amplification cycles by qPCR when compared to uncut samples (Nepon-Sixt and Alexandrow, 2019). Primers specific to the U2OS 2-6-3 reporter gene array amplifying the promoter, 5' β-globin, and 3' β-globin regions were analyzed alongside a gene desert negative control (OR4F5) and active RNA Pol II gene promoter positive control (Pol2-69) to assess locus specific susceptibility to DNase I digestion in WT mU2OS cells. The results are displayed as the percentage of remaining DNase digested chromatin compared to uncut DNA. The active promoter control, which should be the most accessible region analyzed do its constitutively active, euchromatic chromatin state, had just 6.75% DNA remaining following DNase digestion, indicating this region is highly accessible (Figure 23b). In contrast, the negative gene desert control had 28.76% DNA remaining, indicating a more condensed, less accessible chromatin region. Along the mU2OS 2-6-3 gene, the promoter

region was the most accessible (14.92% DNA remaining) despite reporter gene transcription being off, but it was not as accessible as the constitutively active control. The 5' β-globin region had 23.89% DNA remaining, similar to the condensed gene desert negative control and indicative of the transcriptionally inactive reporter gene status (Figure 23b). The more accessible state of the 2-6-3 reporter region despite its heterochromatic status may point to the selectively activatable design of the reporter gene array and supports the idea that gene promoters, regardless of transcription status, are more generally accessible than other gene regions.

Once the baseline effect of DNase I digestion on chromatin accessibility in WT mU2OS cells was established, I used this as a tool to determine if SON depletion results in altered chromatin accessibility at these gene loci. If SON depletion increases locus-specific chromatin accessibility, this would be identified by an increase in the extent of DNase I digestion in SON-depleted cells, yielding less substrate for qPCR, increased cycles necessary for amplification, and a decrease in percent remaining DNA compared to uncut SON-depleted chromatin. To determine the change in accessibility in SON-depleted cells, the percent input chromatin remaining in the siSON condition was divided by the percent input chromatin remaining in the siControl condition. For example at the constitutively active promoter control, SON-depleted chromatin had about one-third (0.37) the percent input DNA remaining compared to control mU2OS chromatin (Figure 24). This decrease in recovery indicates an increase in chromatin accessibility at this active promoter locus when SON is depleted. The gene desert control was more resistant to SON-depleted chromatin accessibility changes (58% recovery). The 2-6-3 promoter region was particularly susceptible to SON depletion (36% recovery), and the 5' β-globin

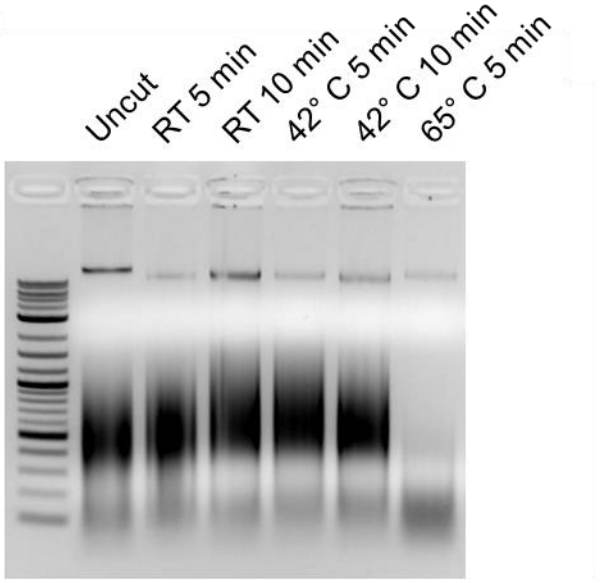
gene showed 44% recovery. KAT5, the promoter to the histone acetyltransferase TIP60, was also assessed for chromatin accessibility changes following SON depletion. A single-replicate SON ChIP-seq previously conducted by our laboratory identified the KAT5 promoter as a region with SON association in HeLa cells (ChIP-seq peak in Figure 24b). In mU2OS cells, SON depletion resulted in increased accessibility of the KAT5 promoter (Figure 24a; 38% recovery). The overall increase in chromatin accessibility observed at all gene loci tested following SON depletion suggests a more global change in chromatin structure that may not be as site-specific as regions where SON association is documented via ChIP.

<u>Figure 23:</u> Regions of highly accessible chromatin are more susceptible to DNase I digestion.

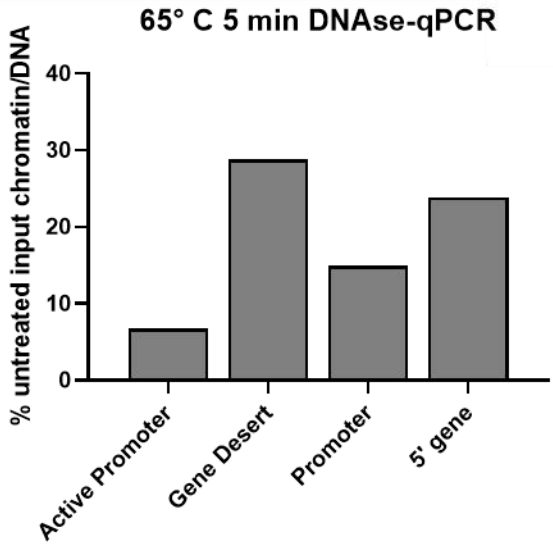
A. WT mU2OS nuclei with equal amounts of input chromatin were either left untreated (lane 1) or treated with DNase I for varying reaction times and temperatures. After digestion, chromatin was mildly sonicated, treated with RNAse A and proteinase K, and DNA purified by PCIA extraction before running on a 1.5% agarose gel. Left lane is a 2-log DNA ladder. Upper band in each lane is the unsonicated portion of the sample, and the lower smear is the sonicated DNA fragments. DNase I treatment for 65° C for 5 min (lane 6) shows marked reduction in total DNA and increased sonication efficiency as indicated by the smaller fragments in the lower smear. White bands across the gel are indicator dyes xylene cyanol (top) and bromophenol blue (bottom).

B. WT mU2OS nuclei were isolated and input split evenly between untreated chromatin and chromatin treated with DNase I for 5 min at 65° C. Samples were sonicated, RNAse/proteinase treated, and DNA purified by PCIA extraction. Equal volume of DNA was loaded in a 96-well dish with 2x SYBER Supermix, nuclease-free water, and primer pairs. Primers amplified a constitutively active RNA Pol II promoter, OR4F5 gene desert, the 2-6-3 CMV promoter, or the 2-6-3 5' β-globin gene regions. Percent input chromatin was calculated using the formula:
$100 \div 2^{Ct_{(DNAse)} - Ct_{(no\ DNAse)}}$ which normalizes the data to uncut input DNA. Ct_{sample} represents the average of 2 technical replicates.

A


Uncut
RT 5 min
RT 10 min
42° C 5 min
42° C 10 min
65° C 5 min

B

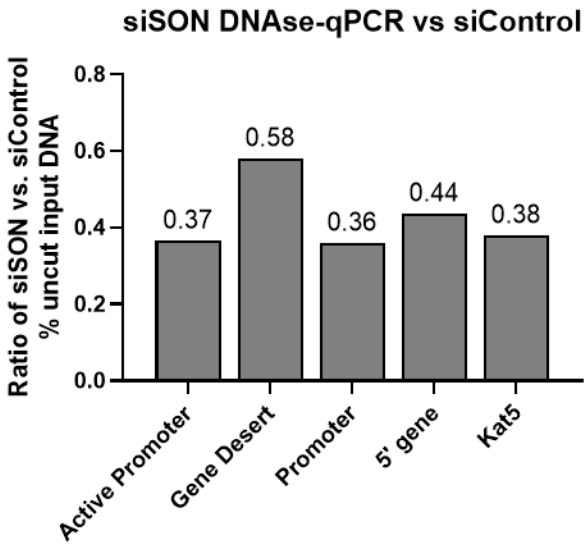

65° C 5 min DNAse-qPCR
% untreated input chromatin/DNA
40
30
20
10
0
Active Promoter
Gene Desert
Promoter
5' gene

<u>Figure 24:</u> SON-depleted gene loci are more susceptible to DNase I digestion.

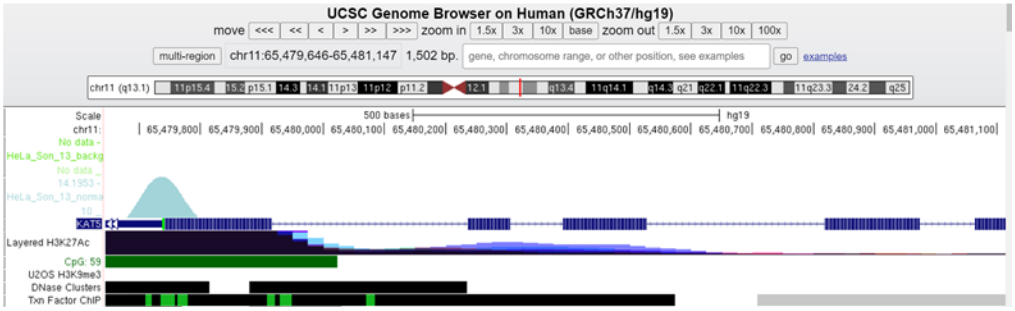
A. mU2OS cells treated with either non-targeting siRNA duplexes or SON si1 + si4 siRNA duplexes were harvested simultaneously and equally split as either untreated or DNase treated at 65° C for 5 min as described in Figure 23. Percent input chromatin was calculated for both siControl and siSON samples, then the siSON percent input divided by the siControl percent input to generate the ratio of chromatin recovered from SON-depleted chromatin for each gene loci. The complete formula: $\frac{(100 \div 2^{Ct_{(DNAse)} - Ct_{(no\ DNAse)}})_{siSON}}{(100 \div 2^{Ct_{(DNAse)} - Ct_{(no\ DNAse)}})_{siControl}}$ where Ct_{sample} represents the average of 2 technical replicates.

B. SON ChIP-seq was performed as a single replicate by a previous Masters student in HeLa cells using anti-SON WU13 antibody. Normalized SON peak at the KAT5 promoter was visualized using the UCSC Genome Browser (GRCh37/hg19) alongside tracks displaying H3K27ac association, CpG islands, H3K9me3 association, DNase susceptibility regions, and transcription factor association at base pair resolution.

A


siSON DNAse-qPCR vs siControl
Ratio of siSON vs. siControl
% uncut input DNA
0.8
0.6
0.4
0.2
0.0
0.37
0.58
0.36
0.44
0.38
Active Promoter
Gene Desert
Promoter
5' gene
Kat5

B


UCSC Genome Browser on Human (GRCh37/hg19)
move
zoom in
zoom out
multi-region
chr11:65,479,646-65,481,147 1,502 bp.
gene, chromosome range, or other position, see examples
go
examples
Scale
500 bases
hg19
chr11:
65,479,800
65,479,900
65,480,000
65,480,100
65,480,200
65,480,300
65,480,400
65,480,500
65,480,600
65,480,700
65,480,800
65,480,900
65,481,000
65,481,100
KAT5
Layered H3K27Ac
CpG: 59
U2OS H3K9me3
DNase Clusters
Txn Factor ChIP

SON depletion alters chromatin organization globally.

Thus far, I have demonstrated that SON associates with an inactive gene promoter (Aim 1), SON depletion is sufficient for chromatin decondensation of a heterochromatic gene locus (Aim 2a), epigenetic marker localization is altered when SON is depleted (Aim 2b), and that SON-depleted chromatin is more accessible via increased susceptibility to DNase I (Aim 2c). The higher order chromatin structure changes observed at the SON-depleted mU2OS locus and the general increase in chromatin accessibility following SON depletion led us to investigate SON's role in global chromatin organization.

To test if SON could be involved in chromatin organization genome-wide, I used a previously developed image processing technique to quantify the extent chromatin condensation in a given nuclei by DAPI staining (Irianto et al., 2014). DAPI is a fluorescent strain that binds to adenine-thymine rich DNA. Regions of the nucleus that have more condensed DNA appear brighter by microscopy, whereas DNA-poor nuclear regions appear darker. For example, nucleoli can be identified as the large, circular, dark spots in a DAPI-stained nucleus. Nuclear speckles, located in gene-poor regions of the nucleus, also correspond to dark spots via DAPI staining. In this way, highly condensed chromatin can also be identified as especially bright areas of the nucleus, such as the areas surrounding nucleoli and the nuclear periphery which is particularly enriched in condensed heterochromatin.

Irianto et al. (2014) developed a technique for identifying highly condensed chromatin regions using Sobel edge detection on DAPI-stained nuclei to calculate a Chromatin Condensation Parameter (CCP) for each nucleus. We adapted the image

processing code first created in Irianto et al. (2014), with few adjustments made to the code to standardize input images and update the Sobel edge detection step due to MATLAB software updates. This code was developed in collaboration with undergraduate researcher Joshua Ward who adapted and ran the code, while I conceived the project and analyzed the resulting data. Example images moving through the processing protocol are displayed in Figure 25 (Full processing protocol described in Methods; MATLAB code available in Appendix B).

As proof of concept, U2OS 2-6-3 cells expressing CFP-LacI and stained with DAPI were imaged to confirm that DAPI staining and Sobel edge detection can recognize chromatin changes associated with reporter gene activation/decondensation (Figure 26). Indeed, the inactive locus was observed as a somewhat brighter DAPI dot (Figure 26a panels a-c), and Sobel edge detection identified the inactive locus as a condensed circle (Figure 26a; panels d-e). The active locus corresponded with a DAPI-depleted nucleus region (Figure 26b; panels a-c), and Sobel edge detection was mostly outside the decondensed locus area (Figure 26b; panels d-e), indicative of the decondensed chromatin structure of the activated reporter locus. With this data, we identified that chromatin structure changes can be detected by DAPI analysis in U2OS 2-6-3 cells.

To ensure our code could detect global changes in chromatin condensation between different cell conditions, we first compared mU2OS cells with α-amanitin treatment to mU2OS cells left untreated. The α-amanitin experiment reported in Figures 16-18 was repurposed for CCP analysis by imaging different cells from these same slides, this time in the DAPI channel. Previous work has demonstrated that the RNA Pol II inhibitor α-amanitin correlates with global changes in chromatin structure, as cells treated

with α-amanitin showed increased chromatin movement and increased chromatin decondensation (Chu et al., 2017; Haaf and Ward 1996; Nagashima et al., 2019; Zidovska et al., 2013). A global decrease in transcription yields a global decrease in chromatin condensation. We therefore expected mU2OS cells treated with α-amanitin to have a lower average CCP compared to untreated cells. A lower CCP means a smaller amount of strong chromatin edges relative to total cell area, and therefore means less chromatin condensation. A variety of analysis parameters were tested to identify the configuration that would yield the largest CCP separation between α-amanitin treated and untreated cells (Figure 27). We determined a pixel reduction factor of 4 and a signal threshold value of 0.03 were optimal for detecting changes in chromatin structure. We found that mU2OS cells treated with α-amanitin had significantly less chromatin condensation compared to untreated mU2OS cells (Figure 28).

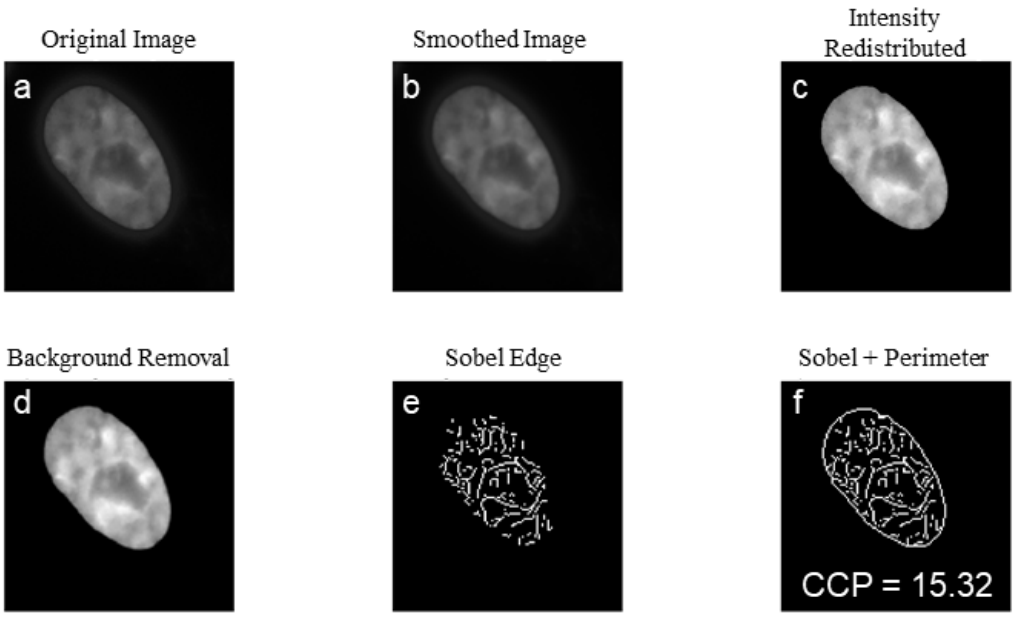
To determine the effect of SON depletion on genome-wide chromatin organization, the control untreated mU2OS cells described above were run alongside SON-depleted nuclei and CCPs calculated. Strikingly, we found that SON-depleted nuclei have significantly increased chromatin condensation compared to control cells (Representative images in Figure 29a; analysis in Figure 29b). This increase in chromatin condensation is not due to any change in nucleus area, as SON-depleted cells exhibit no significant change in nucleus area compared to control cells (Figure 29c). An increase in chromatin condensation indicates more restrictive chromatin movement globally, further implicating SON in chromatin organization.

The complete dataset comparing mU2OS reporter gene activation status and α-amanitin treatment in Control vs. SON-depleted nuclei is displayed in Figure 30, with

statistical analysis in Table 3. Notably, treating SON-depleted cells with α-amanitin results in a global decrease in chromatin condensation similar to control cells treated with α-amanitin (Figure 30 columns 3 vs 4). mU2OS cells transfected with TetON + dox treatment also showed an increase in chromatin condensation compared to control inactive cells that decreased with α-amanitin treatment, indicative of the huge increase in reporter gene transcription activation resulting in more restrictive chromatin movement globally (Figure 30 columns 1, 5 and 6). Interestingly, SON-depleted cells did not show the same increase in chromatin condensation when the reporter gene was activated (Figure 30 columns 3 and 7) as observed in control cells.

<u>Figure 25:</u> Example Chromatin Condensation Parameter Analysis

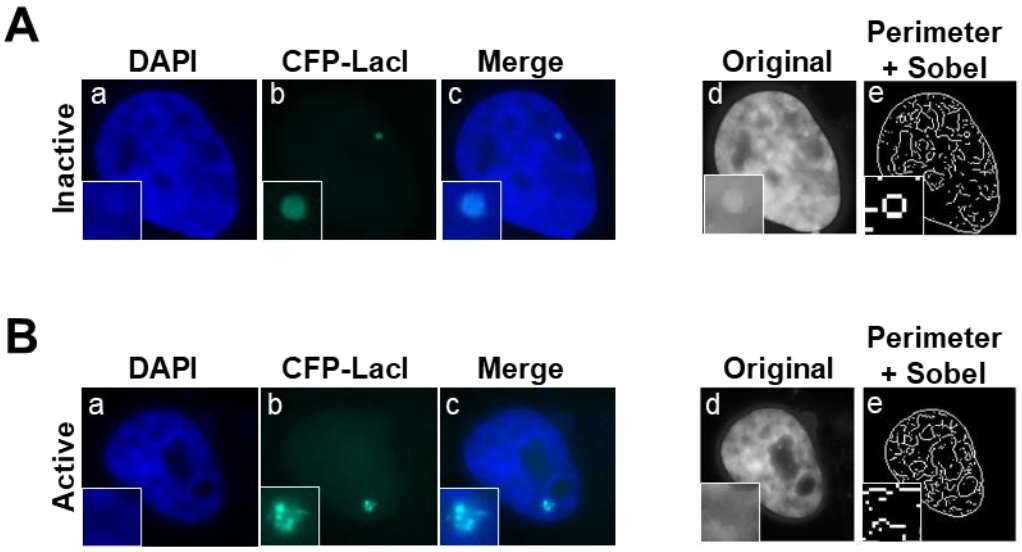
A given DAPI-stained nucleus image was imported into MATLAB (panel a) and iteratively smoothed (panel b), followed by signal thresholding to determine the cell border. Pixel intensity redistribution was used to standardize the image (panel c), and the nucleus was placed on a black background (panel d). Sobel edge detection was applied and holes were filled with black background. The cell perimeter was removed and the inner nucleus placed on a black background (panel e). CCP was calculated by taking the number of pixels identified in the Sobel edge detection vector divided by the total pixel area of the nucleus (panel f). Code adapted from Irianto et al. (2014). Full code available in Appendix B.


Original Image
a
Smoothed Image
b
Intensity Redistributed
c
Background Removal
d
Sobel Edge
e
Sobel + Perimeter
f
CCP = 15.32

<u>Figure 26:</u> U2OS 2-6-3 reporter locus can be identified by Sobel edge detection

A. U2OS 2-6-3 cells were transfected with CFP-LacI and left uninduced, formaldehyde fixed, and stained with DAPI (panels a-c). The DAPI-stained nucleus 8-bit TIF image was run through the CCP pipeline to generate a Sobel edge detected image (panels d-e).

B. U2OS 2-6-3 cells were transfected with CFP-LacI and TetON. 3 hrs post-transfection, dox was added for 2.5 hrs to activate the reporter locus. Cells were formaldehyde fixed and stained with DAPI (panels a-c). The DAPI stained nucleus 8-bit TIF image was run through the CCP pipeline to generate a Sobel edge detected image (panels d-e).

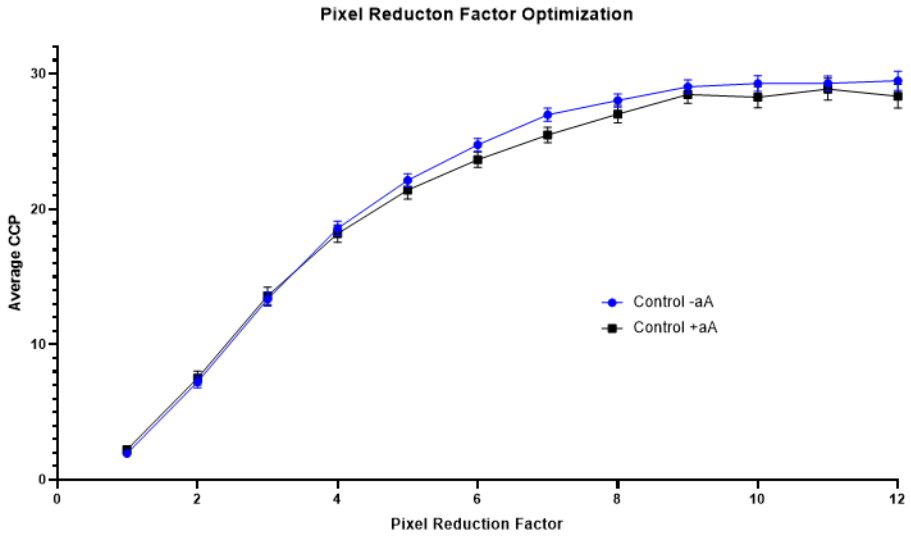

A
DAPI
CFP-LacI
Merge
Original
Perimeter + Sobel
Inactive
a
b
c
d
e
B
DAPI
CFP-LacI
Merge
Original
Perimeter + Sobel
Active
a
b
c
d
e

<u>Figure 27:</u> Chromatin Condensation Parameter Analysis Parameters

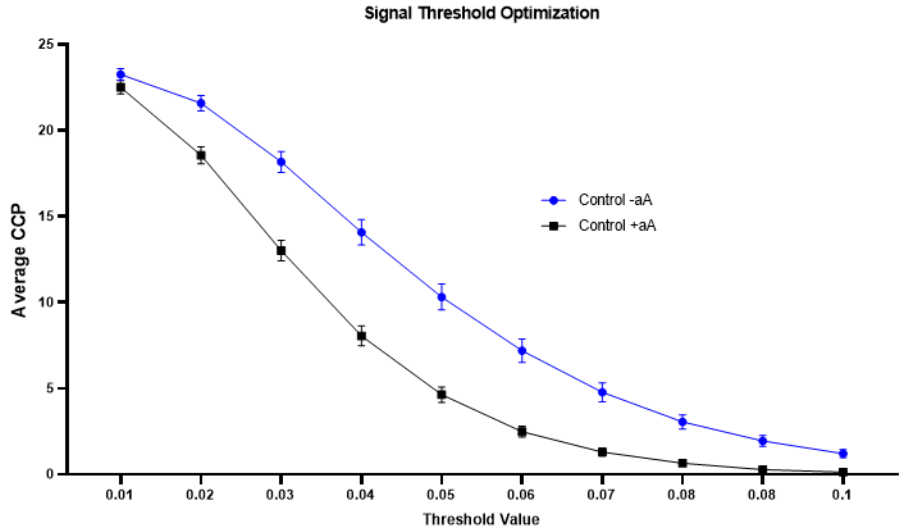
A. Control inactive -aA (n = 41) and control inactive +aA (n = 25) nuclei imaged separately from the same experimental slides in Figures 16-18 were run through the CCP pipeline at varying pixel reduction factors. An increase in pixel reduction factor decreases the total amount of pixels that comprise an image and is necessary for Sobel edge detection. Graph depicts average CCP for each treatment condition vs. pixel reduction factor Bars denote SEM. A pixel reduction factor of 4 was determined as optimal because higher reduction factors resulted in more loss of chromatin resolution via Sobel edge detection.

B. Control inactive -aA (n = 41) and control inactive +aA (n = 25) nuclei imaged separately from the same experimental slides in Figures 16-18 were run through the CCP pipeline with a pixel reduction factor of 4 and varying signal threshold values. A higher signal threshold means less pixels are recognized as strong Sobel edges, decreasing overall CCP. Graph depicts average CCP for each treatment condition vs. threshold value. Bars denote SEM. A signal threshold value of 0.03 was determined as optimal because it yields a large difference in average CCP between treatment conditions while preserving the most pixels for Sobel analysis.

A


Pixel Reducton Factor Optimization
Average CCP
Pixel Reduction Factor
Control -aA
Control +aA
0
10
20
30
0
2
4
6
8
10
12

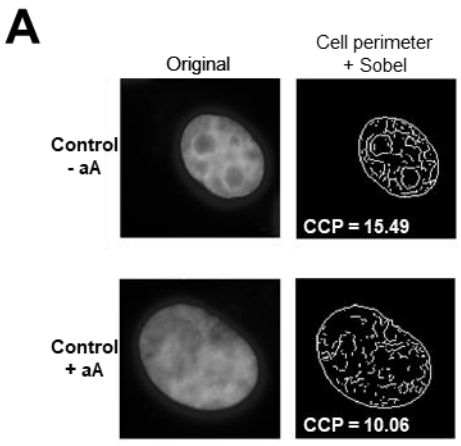
B

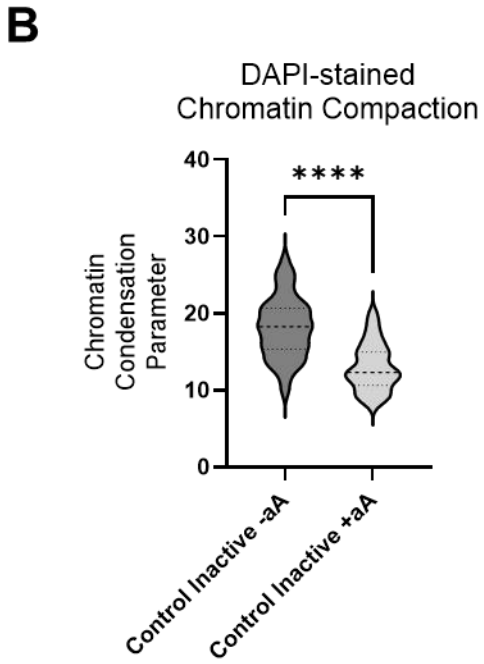

Signal Threshold Optimization
Average CCP
Threshold Value
Control -aA
Control +aA
0
5
10
15
20
25
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.08
0.1

<u>Figure 28:</u> α-Amanitin treatment decreases Chromatin Condensation Parameter

A. Representative control inactive cells without aA (upper) or with aA (lower) treatment. Left panel shows the input image, the right panel shows the image after running through the pipeline and CCP calculated.

B. Control inactive -aA (n = 41) and control inactive +aA (n = 25) nuclei imaged separately from the same experimental slides in Figures 16-18 were run through the image processing pipeline and CCPs calculated for each nucleus. Violin plot depicting CCP distribution in each condition. **** p-value < 0.0001 calculated by three-way ANOVA followed by Tukey's multiple comparisons test on the full data set presented in Figure 30.


A
Original
Cell perimeter
+ Sobel
Control
- aA
CCP = 15.49
Control
+ aA
CCP = 10.06

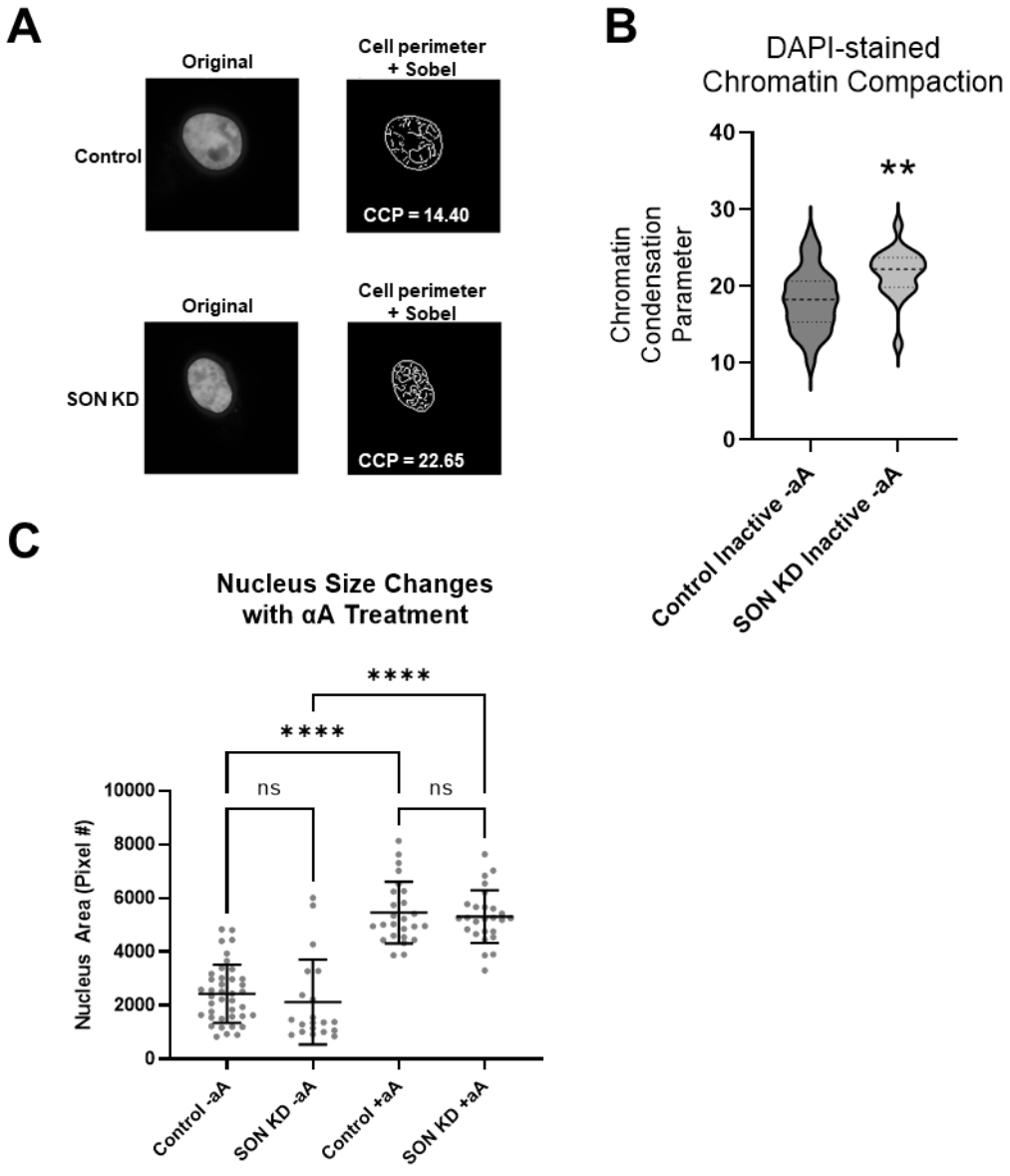

B
DAPI-stained
Chromatin Compaction
Chromatin
Condensation
Parameter
40
30
20
10
0
****
Control Inactive -aA
Control Inactive +aA

<u>Figure 29:</u> SON-depleted nuclei have increased chromatin condensation.

A. Representative control inactive cells without aA (upper) or SON-depleted inactive cells without aA (lower) treatment. Left panel shows the input image, the right panel shows the image after running through the pipeline and CCP calculated.

B. Control inactive -aA (n = 41) and SON KD -aA (n = 20) nuclei imaged separately from the same experimental slides in Figures 16-18 were run through the image processing pipeline and CCPs calculated for each nucleus. Violin plot depicting CCP distribution in each condition. ** p-value < 0.005 calculated by three-way ANOVA followed by Tukey's multiple comparisons test on the full data set presented in Figure 30.

C. Nucleus area in pixels was calculated by the image processing pipeline for mU2OS cells in the Control Inactive -aA, SON KD Inactive -aA, Control Inactive +aA, and SON KD Inactive +aA conditions. Bars denote mean and standard deviation. **** p-value < 0.0001 calculated by three-way ANOVA followed by Tukey's multiple comparisons test.


A
Original
Cell perimeter + Sobel
Control
CCP = 14.40
Original
Cell perimeter + Sobel
SON KD
CCP = 22.65
B
DAPI-stained Chromatin Compaction
Chromatin Condensation Parameter
40
30
20
10
0
**
Control Inactive -aA
SON KD Inactive -aA
C
Nucleus Size Changes with αA Treatment
****
****
ns
ns
Nucleus Area (Pixel #)
10000
8000
6000
4000
2000
0
Control -aA
SON KD -aA
Control +aA
SON KD +aA

<u>Figure 30:</u> Compiled CCP data with pairwise comparisons from Figures 27-29.

These data are identical to the data displayed in Figures 27-29, compiled because all 8 cell conditions shown were analyzed together. Violin plots displaying distribution data; bars indicate mean and upper/lower quartiles.

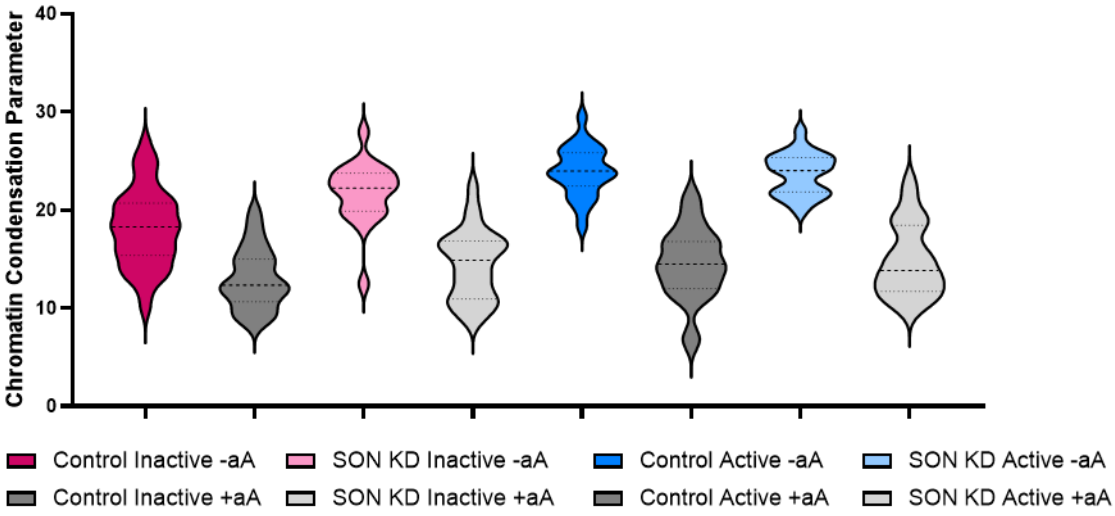

Chromatin Condensation Parameter
40
30
20
10
0
Control Inactive -aA
Control Inactive +aA
SON KD Inactive -aA
SON KD Inactive +aA
Control Active -aA
Control Active +aA
SON KD Active -aA
SON KD Active +aA

Table 3: Tukey's T3 multiple comparisons test of data from Figures 28-30.

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Summary	Adjusted P Value
Control:Inactive -aA vs. Control:Inactive +aA	5.154	2.630 to 7.678	****	<0.0001
Control:Inactive -aA vs. Control:Active -aA	-5.752	-8.343 to -3.161	****	<0.0001
Control:Inactive -aA vs. Control:Active +aA	3.633	1.139 to 6.127	***	0.0004
Control:Inactive -aA vs. SON KD:Inactive -aA	-3.6	-6.311 to -0.8889	**	0.0017
Control:Inactive -aA vs. SON KD:Inactive +aA	3.718	1.224 to 6.212	***	0.0002
Control:Inactive -aA vs. SON KD:Active -aA	-5.499	-8.127 to -2.871	****	<0.0001
Control:Inactive -aA vs. SON KD:Active +aA	3.304	0.6763 to 5.932	**	0.0039
Control:Inactive +aA vs. Control:Active -aA	-10.91	-13.77 to -8.046	****	<0.0001
Control:Inactive +aA vs. Control:Active +aA	-1.522	-4.295 to 1.251	ns	0.6993
Control:Inactive +aA vs. SON KD:Inactive -aA	-8.755	-11.72 to -5.785	****	<0.0001
Control:Inactive +aA vs. SON KD:Inactive +aA	-1.436	-4.209 to 1.337	ns	0.7577
Control:Inactive +aA vs. SON KD:Active -aA	-10.65	-13.55 to -7.759	****	<0.0001
Control:Inactive +aA vs. SON KD:Active +aA	-1.85	-4.744 to 1.044	ns	0.5123
Control:Active -aA vs. Control:Active +aA	9.384	6.551 to 12.22	****	<0.0001
Control:Active -aA vs. SON KD:Inactive -aA	2.152	-0.8751 to 5.179	ns	0.3696
Control:Active -aA vs. SON KD:Inactive +aA	9.47	6.636 to 12.30	****	<0.0001
Control:Active -aA vs. SON KD:Active -aA	0.2531	-2.699 to 3.205	ns	>0.9999
Control:Active -aA vs. SON KD:Active +aA	9.056	6.104 to 12.01	****	<0.0001
Control:Active +aA vs. SON KD:Inactive -aA	-7.233	-10.18 to -4.288	****	<0.0001
Control:Active +aA vs. SON KD:Inactive +aA	0.08561	-2.660 to 2.831	ns	>0.9999
Control:Active +aA vs. SON KD:Active -aA	-9.131	-12.00 to -6.263	****	<0.0001
Control:Active +aA vs. SON KD:Active +aA	-0.3285	-3.196 to 2.539	ns	>0.9999
SON KD:Inactive -aA vs. SON KD:Inactive +aA	7.318	4.374 to 10.26	****	<0.0001
SON KD:Inactive -aA vs. SON KD:Active -aA	-1.899	-4.957 to 1.160	ns	0.5513
SON KD:Inactive -aA vs. SON KD:Active +aA	6.904	3.846 to 9.963	****	<0.0001
SON KD:Inactive +aA vs. SON KD:Active -aA	-9.217	-12.08 to -6.349	****	<0.0001
SON KD:Inactive +aA vs. SON KD:Active +aA	-0.4141	-3.282 to 2.454	ns	0.9998
SON KD:Active -aA vs. SON KD:Active +aA	8.803	5.818 to 11.79	****	<0.0001

CHAPTER VI

DISCUSSION

Nuclear speckles are domains in higher-order eukaryotes that are the storage and assembly sites of pe-mRNA processing factors (Mao et al., 2011; Spector and Lamond, 2011; Tripathi et al., 2012). Phosphorylation regulates the shuttling of these factors between sites of active transcription and speckles (Bernard et al., 2010; Galganski et al., 2017; Tripathi et al., 2010). Previous studies from our laboratory have focused on the large protein SON and its role in nuclear speckle organization, demonstrating that proper speckle morphology requires SON. Loss of SON results in a donut-shaped arrangement of nuclear speckle factors (Sharma et al. 2010). Our laboratory has also shown that SON's unique repetitive motifs are required for nuclear speckle organization and maintaining proper splicing of a host of protein-coding genes including the histone modifier HDAC6 (Battini et al. 2015; Sharma et al., 2011).

In addition to SON's role in nuclear organization and splicing, it has also been implicated in gene expression via DNA binding (Sun et al., 2001). More recently, SON has been detected at gene promoters by ChIP-seq analysis (Kim et al., 2016). This was particularly interesting because neither SON's nuclear speckle localization or association with weak splice sites is known to involve DNA-binding, suggesting a completely new function for SON at chromatin. The larger goal of my thesis was therefore to

determine SON's role in chromatin-mediated gene expression. I used the U2OS 2-6-3 inducible reporter gene array as a tool to map SON's association with the heterochromatic reporter gene locus, and to determine SON's role in chromatin organization both at the reporter locus and genome-wide. A proposed model for SON association with the reporter gene locus is depicted in Figure 31.

SON in gene promoter regulation

SON was first identified as a DNA-binding protein by Mattioni et al. (1992) and Sun et al. (2001), and more recently by Kim et al. (2016) who found that SON associates with gene promoters. Here, we show that SON associates with a transcriptionally silent reporter gene promoter, and SON is not detectable at the transcriptionally active locus. SON's association with the heterochromatic reporter locus further supports a role for SON in gene regulation at promoters. It was previously shown that SON indirectly regulates the transcription factor GATA-2 by suppressing the promoter of GATA-2-targeting miR-27a (Ahn et al., 2013). Here, I show that SON promoter association with the U2OS 2-6-3 reporter gene locus is specific to the transcriptionally inactive state, again implicating SON in promoter repression.

It is still unclear if gene repression is the only function for SON at promoter regions, but preliminary ChIP-seq data in our laboratory suggests that SON associates with the promoters of euchromatic housekeeping genes related to ribosome and RNA processing (unpublished). This may add some nuance to SON's suggested role in promoter repression since housekeeping genes are typically constitutively active. When

Kim et al. (2016) identified SON at gene promoters, they found that SON depletion was correlated with both an increase in H3K4me3 at target genes, and an increase in target gene transcript. It will be interesting to tease out exactly how SON affects different target genes through promoter association.

SON's repeat motifs in gene expression

Previous work in our laboratory has demonstrated that SON repeats are required for accumulation at a reporter BTM minigene (Battini et al., 2013; Sharma et al., 2011), so it is possible SON functions in a similar way with chromatin remodelers at promoters. I have demonstrated here that SON's repeat motifs are necessary for U2OS 2-6-3 reporter locus association. Additionally, SON's C-terminal RNA associating domains that are necessary for splicing regulation (Ahn et al., 2011), are not required for chromatin association. This suggests that SON's chromatin associating function is distinct from its role in splicing.

SON's repeat motifs comprise over one-third of its amino acid sequence and have not been found in any other protein. Given that SON has been identified in maintaining stability and/or organization of cellular processes, we think SON's repeat motifs probably function as a type of molecular scaffold. The most prominent example of a protein scaffold is the C-terminal domain of RNA Pol II (CTD), although there is no sequence similarity to SON. Pre-mRNA processing factors dock at the CTD YSPTSPS repeats which increases efficiency of co-transcriptional processing (Ahn and Buratowski, 2004; Bataille et al., 2012; Lenasi and Barboric, 2013; Misteli and Spector, 1999). Future work

will determine what factors associate with SON and when, and the mechanisms that govern these interactions.

SON in transcription activation

In specific aim 1, I observed that SON does not colocalize with the transcriptionally activated reporter gene locus, and this decrease in colocalization preceded higher order chromatin decondensation and reporter transcript detection by microscopy. This suggests that SON functions in either maintaining the inactive status of the reporter gene, or perhaps functions in keeping the reporter gene in a poised state for recruitment of early transcription initiation machinery. There is evidence for a class of proteins termed "splicing adapters" that recognize and interpret epigenetic markers to modulate splicing events (Agirre at al., 2015; Luco et al., 2011; Matveeva et al., 2019; Solboleva et al., 2017), however it is unlikely that SON is functioning as a splicing adapter at the mU2OS reporter promoter since alternative splicing cannot take place at this gene.

Additionally, I found that SON association with the reporter locus is reversible, and replacement of dox-containing medium to achieve locus transcription inactivation of fully decondensed loci is sufficient for SON colocalization with the inactivated locus. The dynamics of how and when SON associates with the inactivated reporter locus is still unknown. This could be addressed by analyzing SON localization at various timepoints after media replacement. It would be interesting to see if chromatin re-condensation occurs before or after SON recruitment to the inactivated locus. When I inactivated the

reporter gene using the TetOFF system, SON was not particularly enriched at the "TetOFF + dox" inactivated locus, even though most inactivated loci were visibly condensed. I attributed this to the experimental setup, with more time probably being necessary to completely remove tTA from the promoter to inactivate the reporter gene. Perhaps this lack of robust SON association with the condensed TetOFF inactivated locus is an indicator that chromatin re-condensation precedes SON recruitment to the inactivated locus. The dynamics of gene inactivation and the epigenetic factors that contribute to modulation in higher order chromatin structure is still not well known, and further characterization of SON's role in maintaining chromatin stability at inactive promoters will enhance our understanding of the coordination between 3D chromatin structure and gene expression.

SON in epigenetic regulation

There is evidence that transcription factors can recognize and/or manipulate chromatin markers to modulate chromatin accessibility (Muller et al., 2001; Ren et al., 2018) In my experiments I showed the heterochromatic histone marker H3K9me3 is localized to the SON-depleted inactive reporter locus, and that H3K9me3 is colocalized with the transcriptionally active SON-depleted locus but not the control active locus. During the early stages after transcription activation, H3K9me3 is removed from the U2OS 2-6-3 reporter gene locus (Hyun et al., 2017; Janicki et al., 2004). It is interesting that SON-depleted reporter loci are enriched in H3K9me3 even in the transcriptionally active condition. Perhaps SON-depletion inhibits H3K9me3 removal, although the increase in H3K9me3 at SON-depleted loci does not seem to affect transcription factor

recruitment to the SON-depleted active locus. Indeed, SON-depletion has already been linked to an increase in H3K4me3 at SON target genes (Kim et al., 2016), although H3K4me3 does not appear to associate with the mU2OS gene to corroborate these findings. The increase in H3K4me3 at SON-depleted gene targets was correlated with target gene upregulation, further implicating SON in promoter repression (Kim et al., 2016). It will be interesting to further explore the role of SON in epigenetic modulation of target genes and its effect on gene expression.

HP1α, a chromatin modifier that recognizes the methylated H3K9 mark to suppress transcription, shows weak colocalization with the SON-depleted reporter locus despite the increase in H3K9me3 colocalization. However, it is worth noting that I did not observe the robust recruitment of HP1α to control inactive loci as reported in Janicki et al. (2004). This is most likely due to my use of commercially available antibody to assess HP1α localization. The highly condensed, heterochromatic status of the inactive reporter locus is particularly difficult to access by antibody-epitope binding, which then appears as a lack of recruitment to the condensed locus due to accessibility issues. Therefore, it is difficult to say with certainty how or if HP1α localization is altered at the SON-depleted reporter gene locus. In the future, live imaging studies employing fluorescent-tagged protein colocalization would be a more reliable method to assess protein dynamics at the U2OS 2-6-3 locus.

SON and RNA Pol II-mediated transcription elongation

Some inactive genes have RNA Pol II bound to the promoter but not elongating the mRNA, allowing the gene to be transcriptionally stalled and poised for activation when serine 5 of the C-terminal domain is phosphorylated (Gilchrist et al., 2010; Margaritis and Holstege, 2008; Muse et al., 2007). Although I did not specifically assess the Ser5-phosphorylated RNA Pol II, a poised RNA pol II does not seem likely for the mU2OS gene because the silent locus is heterochromatic and transcriptionally naïve (Janicki et al., 2004). To rule out that SON interaction with the mU2OS gene is tied to RNA pol II localization, I assessed RNA pol II colocalization with mCherry-LacI. RNA pol II showed no colocalization with the inactive reporter locus in control or SON depleted cells.

In specific aim 2, we blocked global transcription elongation by α-amanitin treatment to rule out that SON-depleted decondensed reporter loci were not due to some spontaneous reporter gene activation. Our control cells were consistent with established literature, with transcriptionally inactive loci not changing in size when treated with α-amanitin alone, increasing in size with transcription activation alone, and decreasing in size similar to control inactive loci when activated by TetON/dox with α-amanitin treatment. These observations are consistent with previous work in our laboratory (Varia et al., 2013) as well as others (Refalska-Metcalf et al., 2010) who reported pre-treatment with α-amanitin prevents large-scale chromatin decondensation. Cells treated with α-amanitin have acetylated loci and exhibit Brd2 and GCN5 recruitment despite a lack of chromatin decondensation, and loci appear significantly decreased in size compared to transcriptionally active loci (Refalska-Metcalf et al., 2010).

SON in chromatin organization

In specific aim 2, I investigated SON's role in chromatin organization. I used SON depletion studies to determine how SON affects chromatin stability at the reporter gene. I found that SON-depleted reporter loci are decondensed yet transcriptionally silent, and that SON-depleted loci are still activatable. To our knowledge, this is the first study to observe a decondensed gene locus that is uncoupled from transcription. The inducible U2OS 2-6-3 reporter gene array is therefore a valuable tool in isolating these distinct processes, as this dynamic uncoupling is likely not possible at endogenous gene loci. Recent studies support the idea that chromatin organization and transcription activation are distinct processes, but this has proved difficult to investigate in vivo because an increase in chromatin accessibility is correlated with an increase in transcription (Chambeyron and Bickmore, 2004; Hausnerova and Lanctot, 2017). The tightly controlled CMV promoter in the U2OS 2-6-3 reporter gene array does not permit transcription activation without expression of TetON in the presence of dox, allowing us to separate activation from chromatin decondensation.

Notably, my experiments were all conducted in U2OS 2-6-3 cells that expressed a stably incorporated mCherry-LacI for reporter locus labeling (mU2OS cells). A recent study by Hochberg-Laufer et al. (2019) examining SRSF7 dynamics at the transiently expressed RFP-LacI U2OS 2-6-3 locus did not report any changes in locus area following SON depletion. In the original U2OS 2-6-3 cell line, reporter loci are labeled by transient transfection of fluorescently tagged LacI. When I transfected CFP-LacI into U2OS 2-6-3 cells, I observed much higher endogenous and exogenous SON colocalization with CFP-LacI compared to endogenous SON colocalization with wild-type mU2OS cells. It may

be that these mU2OS cells more closely resemble endogenous gene loci due to their more "passive" chromatin state that is not disrupted by transient FP-LacI expression. This works in our favor for studying higher order chromatin structure at the mU2OS gene locus.

The rounded appearance and increased size of SON-depleted mU2OS loci is almost reminiscent of our previous work showing speckle factor reorganization into a donut phenotype following SON depletion (Sharma et al., 2010). While I did find that H3K9me3 localization is altered at SON-depleted mU2OS loci, it would be interesting to dive further into how SON maintains chromatin condensation, and if this is by a similar mechanism to the way it maintains nuclear speckle integrity. SON's unique repeat motifs are required for both speckle organization and chromatin condensation, so perhaps the repeat domain acts as a sort of molecular scaffold for chromatin or other factors to properly assemble.

SON's chromatin association function vs. splicing regulation

A growing body of SON multi-omics data has given new insight into SON's role in gene regulation. Exon array analysis in our laboratory by Sharma et al. (2011) identified SON depletion was associated with both gene activation and repression of target genes involved in cellular processes such as apoptosis, cell cycle regulation, and Wnt signaling, among others. SON depletion also altered the splicing of genes via exon exclusion or inclusion (Sharma et al., 2011). The alternatively spliced genes with the highest statistical probability were involved in histone modification, which is particularly

interesting in the context of my thesis work. Histone de-acetylase 6 (HDAC6) and histone methyltransferase 8 (SetD8) exhibited exon skipping following SON depletion, despite no change in transcript level (Battini et al., 2013; Sharma et al., 2011). While my thesis investigated direct SON association with a reporter gene promoter and its affect on chromatin organization, it is important to consider that SON has a role in several different, distinct cellular processes, and indirect effects of SON depletion on essential genes like chromatin modifiers probably accounts for some of the genome-wide chromatin structure changes I observed.

Thus far, there is no reported overlap in the list of top genes upregulated/downregulated after SON depletion (Sharma et al., 2011), top genes with slicing defects after SON depletion (Sharma et al., 2011), and top genes identified by SON ChIP-seq (Kim et al., 2016). This suggests that SON's role in transcript regulation, pre-mRNA processing, and chromatin association are functionally distinct. It is a major long-term goal to determine how or if these SON-mediated gene expression roles are interconnected.

SON in chromatin accessibility

I also examined SON's role in genome-wide chromatin organization. I found that SON depleted nuclei are more susceptible to DNase digestion and examined how SON depletion affects global chromatin structure by microscopy. Early research has shown that treatment of nuclei with DNase did not abolish SC35-labeled nuclear speckles (Spector et al., 1991). Although DNase maintains speckle integrity, SON-depleted cells

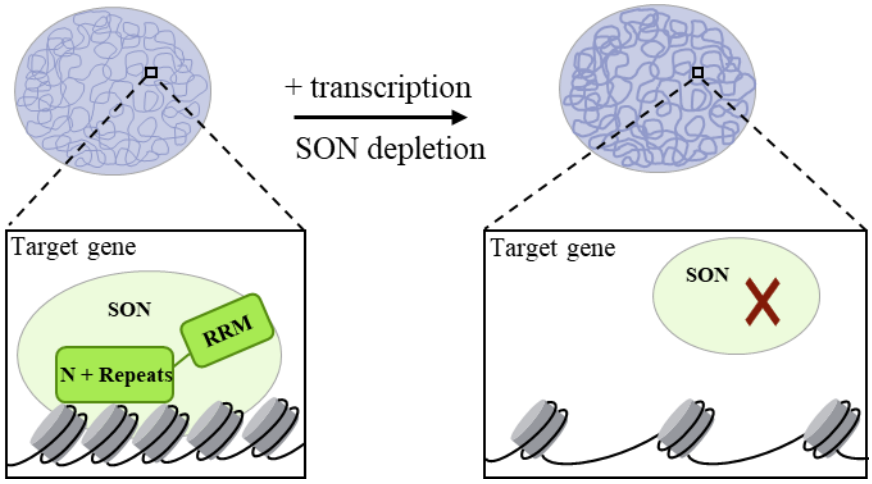
have significantly altered speckle morphology and reorganization of nuclear speckle factors (Sharma et al., 2010). Therefore, we cannot completely rule out that the increased chromatin accessibility I observed at SON-depleted loci could be related to SON's role in nuclear speckle organization. Furthermore, the increased chromatin compaction I observed in SON-depleted DAPI stained nuclei indicates that SON plays an important role in maintaining chromatin stability genome-wide.

Increasingly complex multi-omics methods give valuable insights into protein functions in the context of higher order 3D chromatin structure. Despite its widespread expression, SON has historically been an understudied protein, and this is likely due to a lack of technologies available to reliably study it. With over a third of SON's amino acid sequence consists of tandem repeat motifs, the size, insolubility, and repetitive nature of SON have made it notoriously difficult to work with in a laboratory setting. The explosion of interest in 3D chromatin organization and nuclear speckle structure this past decade has brought on a slew of new next-generation sequencing technologies like Hi-C and TSA-seq, and the rise of super-resolution microscopy can more robustly and reliably study complex chromatin structures and nuclear speckles. With these advancements have come a renewed interest in SON and its role in nuclear organization and gene expression. I am particularly interested in assay for transposable accessible chromatin (ATAC-seq), which maps regions of accessible chromatin in a given cell population. Comparing control vs. SON-depleted cells by ATAC-seq would give new insights into SON's role in chromatin accessibility, and is the next logical step based on my thesis work presented here. I have demonstrated that SON associates with a reporter gene promoter and that SON depletion results in chromatin decondensation that is more susceptible to DNase

digestion. Furthermore, SON depleted nuclei have an overall increased chromatin condensation parameter (CCP) suggesting that chromatin structure genome-wide is altered when SON is depleted. Mapping these more accessible regions by ATAC-seq would give further insight into SON's role in chromatin accessibility at specific gene loci. I would expect ATAC-seq results to be consistent with the target genes identified by SON ChIP-seq. Significant overlap in these gene profiles would support the idea that SON occupancy at gene promoters is directly correlated with increased chromatin stability, and that SON depleted promoters exhibit increased chromatin accessibility. Overall, my thesis work fits well into the emerging model that 3D chromatin structure is an important regulatory mechanism for maintaining proper gene expression. SON's established role in nuclear speckle organization and splicing regulation, along with my own findings that SON is involved in promoter gene repression and chromatin accessibility, point to a broader function of SON in nuclear stability.

Figure 31: Model for SON's role in chromatin stability

SON repeat motifs associate with target gene promoters to maintain chromatin stability and repress transcription. When SON is not occupying promoters, either by target gene activation or SON depletion, target genes are more decondensed and accessible.



REFERENCES

4D Nucleome. (2022, February 2). *Program Snapshot.* https://commonfund.nih.gov/4dnucleome

Agirre, E., Bellora, N., Alló, M., Pagès, A., Bertucci, P., Kornblihtt, A. R., & Eyras, E. (2015). A chromatin code for alternative splicing involving a putative association between CTCF and HP1α proteins. *BMC biology*, *13*, 31.

Ahn, E. Y., DeKelver, R. C., Lo, M. C., Nguyen, T. A., Matsuura, S., Boyapati, A., Pandit, S., Fu, X. D., & Zhang, D. E. (2011). SON controls cell-cycle progression by coordinated regulation of RNA splicing. *Molecular cell*, *42*(2), 185–198.

Ahn, E. E., Higashi, T., Yan, M., Matsuura, S., Hickey, C. J., Lo, M. C., Shia, W. J., DeKelver, R. C., & Zhang, D. E. (2013). SON protein regulates GATA-2 through transcriptional control of the microRNA 23a~27a~24-2 cluster. *The Journal of biological chemistry*, *288*(8), 5381–5388.

Ahn, S. H., Kim, M., & Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Molecular cell*, *13*(1), 67–76.

Alberti, S., Gladfelter, A., & Mittag, T. (2019). Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell*, *176*(3), 419–434.

Anania, C., & Lupiáñez, D. G. (2020). Order and disorder: abnormal 3D chromatin organization in human disease. *Briefings in functional genomics*, *19*(2), 128–138.

Banani, S. F., Lee, H. O., Hyman, A. A., & Rosen, M. K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nature reviews. Molecular cell biology*, *18*(5), 285–298.

Bataille, A. R., Jeronimo, C., Jacques, P. É., Laramée, L., Fortin, M. È., Forest, A., Bergeron, M., Hanes, S. D., & Robert, F. (2012). A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Molecular cell*, *45*(2), 158–170.

Battini, V. P., Bubulya, A., & Bubulya, P. A. (2015). Accurate splicing of HDAC6 pre-mRNA requires SON. *International journal of molecular sciences*, *16*(3), 5886–5899.

Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., & Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods (San Diego, Calif.)*, *58*(3), 268–276.

Bensaude O. (2011). Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity?. *Transcription*, *2*(3), 103–108.

Bernard, D., Prasanth, K. V., Tripathi, V., Colasse, S., Nakamura, T., Xuan, Z., Zhang, M. Q., Sedel, F., Jourdren, L., Coulpier, F., Triller, A., Spector, D. L., & Bessis, A. (2010). A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *The EMBO journal*, *29*(18), 3082–3093.

Boeynaems, S., Alberti, S., Fawzi, N. L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L., Tompa, P., & Fuxreiter, M. (2018). Protein Phase Separation: A New Phase in Cell Biology. *Trends in cell biology*, *28*(6), 420–435.

Boija, A., Klein, I. A., Sabari, B. R., Dall'Agnese, A., Coffey, E. L., Zamudio, A. V., Li, C. H., Shrinivas, K., Manteiga, J. C., Hannett, N. M., Abraham, B. J., Afeyan, L. K., Guo, Y. E., Rimel, J. K., Fant, C. B., Schuijers, J., Lee, T. I., Taatjes, D. J., & Young, R. A. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*, *175*(7), 1842–1855.e16.

Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, Jülicher F, Hyman AA. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science, 324*(5935), 1729-32.

Brody, Y., Neufeld, N., Bieberstein, N., Causse, S. Z., Böhnlein, E. M., Neugebauer, K. M., Darzacq, X., & Shav-Tal, Y. (2011). The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS biology*, *9*(1), e1000573.

Brown, A. L., Wilkins, O. G., Keuss, M. J., Hill, S. E., Zanovello, M., Lee, W. C., Bampton, A., Lee, F., Masino, L., Qi, Y. A., Bryce-Smith, S., Gatt, A., Hallegger, M., Fagegaltier, D., Phatnani, H., NYGC ALS Consortium, Newcombe, J.,

Gustavsson, E. K., Seddighi, S., Reyes, J. F., … Fratta, P. (2022). TDP-43 loss and ALS-risk SNPs drive mis-splicing and depletion of UNC13A. *Nature*, *603*(7899), 131–137.

Brown, J. M., Green, J., das Neves, R. P., Wallace, H. A., Smith, A. J., Hughes, J., Gray, N., Taylor, S., Wood, W. G., Higgs, D. R., Iborra, F. J., & Buckle, V. J. (2008). Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *The Journal of cell biology*, *182*(6), 1083–1097.

Bubulya, P. A., Prasanth, K. V., Deerinck, T. J., Gerlich, D., Beaudouin, J., Ellisman, M. H., Ellenberg, J., & Spector, D. L. (2004). Hypophosphorylated SR splicing factors transiently localize around active nucleolar organizing regions in telophase daughter nuclei. *The Journal of cell biology*, *167*(1), 51–63.

Callan H. G. (1986). Lampbrush chromosomes. *Molecular biology, biochemistry, and biophysics*, *36*, 1–252.

Chambeyron, S., & Bickmore, W. A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes & development*, *18*(10), 1119–1130.

Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E. K., Adam, S. A., Goldman, R., van Steensel, B., Ma, J., & Belmont, A. S. (2018). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *The Journal of cell biology*, *217*(11), 4025–4048.

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G. M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., & Tjian, R. (2018). Imaging dynamic and

selective low-complexity domain interactions that control gene transcription. *Science (New York, N.Y.), 361*(6400), eaar2555.

Chu, F. Y., Haley, S. C., & Zidovska, A. (2017). On the origin of shape fluctuations of the cell nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(39), 10338–10343.

Chung, I. M., Ketharnathan, S., Kim, S. H., Thiruvengadam, M., Rani, M. K., & Rajakumar, G. (2016). Making Sense of the Tangle: Insights into Chromatin Folding and Gene Regulation. *Genes*, *7*(10), 71.

Cremer, T., & Cremer, M. (2010). Chromosome territories. *Cold Spring Harbor perspectives in biology*, *2*(3), a003889.

Dekker, J., & Heard, E. (2015). Structural and functional diversity of Topologically Associating Domains. *FEBS letters*, *589*(20 Pt A), 2877–2884.

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science (New York, N.Y.)*, *295*(5558), 1306–1311.

Fei, J., Jadaliha, M., Harmon, T. S., Li, I., Hua, B., Hao, Q., Holehouse, A. S., Reyer, M., Sun, Q., Freier, S. M., Pappu, R. V., Prasanth, K. V., & Ha, T. (2017). Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. *Journal of cell science, 130*(24), 4180–4192.

Feric, M., Vaidya, N., Harmon, T. S., Mitrea, D. M., Zhu, L., Richardson, T. M., Kriwacki, R. W., Pappu, R. V., & Brangwynne, C. P. (2016). Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell*, *165*(7), 1686–1697.

Furukawa, T., Tanji, E., Kuboki, Y., Hatori, T., Yamamoto, M., Shimizu, K., Shibata, N., & Shiratori, K. (2012). Targeting of MAPK-associated molecules identifies SON as a prime target to attenuate the proliferation and tumorigenicity of pancreatic cancer cells. *Molecular cancer*, *11*, 88.

Galganski, L., Urbanek, M. O., & Krzyzosiak, W. J. (2017). Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic acids research*, *45*(18), 10350–10368.

Gilchrist, D. A., Dos Santos, G., Fargo, D. C., Xie, B., Gao, Y., Li, L., & Adelman, K. (2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell*, *143*(4), 540–551.

Gröschel, S., Sanders, M. A., Hoogenboezem, R., de Wit, E., Bouwman, B., Erpelinck, C., van der Velden, V., Havermans, M., Avellino, R., van Lom, K., Rombouts, E. J., van Duin, M., Döhner, K., Beverloo, H. B., Bradner, J. E., Döhner, H., Löwenberg, B., Valk, P., Bindels, E., de Laat, W., … Delwel, R. (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell*, *157*(2), 369–381.

Haaf, T., & Ward, D. C. (1996). Inhibition of RNA polymerase II transcription causes chromatin decondensation, loss of nucleolar structure, and dispersion of chromosomal domains. *Experimental cell research*, *224*(1), 163–173.

Hennig, S., Kong, G., Mannen, T., Sadowska, A., Kobelke, S., Blythe, A., Knott, G. J., Iyer, K. S., Ho, D., Newcombe, E. A., Hosoki, K., Goshima, N., Kawaguchi, T., Hatters, D., Trinkle-Mulcahy, L., Hirose, T., Bond, C. S., & Fox, A. H. (2015).

Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *The Journal of cell biology*, *210*(4), 529–539.

Hickey, C. J., Kim, J. H., & Ahn, E. Y. (2014). New discoveries of old SON: a link between RNA splicing and cancer. *Journal of cellular biochemistry*, *115*(2), 224–231.

Hochberg-Laufer, H., Neufeld, N., Brody, Y., Nadav-Eliyahu, S., Ben-Yishay, R., & Shav-Tal, Y. (2019). Availability of splicing factors in the nucleoplasm can regulate the release of mRNA from the gene after transcription. *PLoS genetics*, *15*(11), e1008459.

Hoffman, B. G., Zavaglia, B., Witzsche, J., Ruiz de Algara, T., Beach, M., Hoodless, P. A., Jones, S. J., Marra, M. A., & Helgason, C. D. (2008). Identification of transcripts with enriched expression in the developing and adult pancreas. *Genome biology*, *9*(6), R99.

Hondele M, Sachdev R, Heinrich S, Wang J, Vallotton P, Fontoura BMA, Weis K. (2019). DEAD-box ATPases are global regulators of phase-separated organelles. *Nature*, *573*(7772), 144-148.

Hu, Y., Plutz, M., & Belmont, A. S. (2010). Hsp70 gene association with nuclear speckles is Hsp70 promoter specific. *The Journal of cell biology*, *191*(4), 711–719.

Hu, S., Lv, P., Yan, Z., & Wen, B. (2019). Disruption of nuclear speckles reduces chromatin interactions in active compartments. *Epigenetics & chromatin*, *12*(1), 43.

Hübner, M. R., Eckersley-Maslin, M. A., & Spector, D. L. (2013). Chromatin organization and transcriptional regulation. *Current opinion in genetics & development*, *23*(2), 89–95.

Hyun, K., Jeon, J., Park, K., & Kim, J. (2017). Writing, erasing and reading histone lysine methylations. *Experimental & molecular medicine*, *49*(4), e324.

Ilik, İ. A., Malszycki, M., Lübke, A. K., Schade, C., Meierhofer, D., & Aktaş, T. (2020). SON and SRRM2 are essential for nuclear speckle formation. *eLife*, *9*, e60579.

Irianto, J., Lee, D. A., & Knight, M. M. (2014). Quantification of chromatin condensation level by image processing. *Medical engineering & physics*, *36*(3), 412–417.

Janicki, S. M., Tsukamoto, T., Salghetti, S. E., Tansey, W. P., Sachidanandam, R., Prasanth, K. V., Ried, T., Shav-Tal, Y., Bertrand, E., Singer, R. H., & Spector, D. L. (2004). From silencing to gene expression: real-time analysis in single cells. *Cell*, *116*(5), 683–698.

Karlas, A., Machuy, N., Shin, Y., Pleissner, K. P., Artarini, A., Heuer, D., Becker, D., Khalil, H., Ogilvie, L. A., Hess, S., Mäurer, A. P., Müller, E., Wolff, T., Rudel, T., & Meyer, T. F. (2010). Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature*, *463*(7282), 818–822.

Kim, J. H., Baddoo, M. C., Park, E. Y., Stone, J. K., Park, H., Butler, T. W., Huang, G., Yan, X., Pauli-Behn, F., Myers, R. M., Tan, M., Flemington, E. K., Lim, S. T., & Ahn, E. Y. (2016). SON and Its Alternatively Spliced Isoforms Control MLL Complex-Mediated H3K4me3 and Transcription of Leukemia-Associated Genes. *Molecular cell*, *61*(6), 859–873.

Kim, J. H., Jeong, K., Li, J., Murphy, J. M., Vukadin, L., Stone, J. K., Richard, A., Tran, J., Gillespie, G. Y., Flemington, E. K., Sobol, R. W., Lim, S. S., & Ahn, E. E. (2021). SON drives oncogenic RNA splicing in glioblastoma by regulating PTBP1/PTBP2 switching and RBFOX2 activity. *Nature communications*, *12*(1), 5551.

Kim, J. H., Shinde, D. N., Reijnders, M., Hauser, N. S., Belmonte, R. L., Wilson, G. R., Bosch, D., Bubulya, P. A., Shashi, V., Petrovski, S., Stone, J. K., Park, E. Y., Veltman, J. A., Sinnema, M., Stumpel, C., Draaisma, J. M., Nicolai, J., University of Washington Center for Mendelian Genomics, Yntema, H. G., Lindstrom, K., … Ahn, E. (2016). De Novo Mutations in SON Disrupt RNA Splicing of Genes Essential for Brain Development and Metabolism, Causing an Intellectual-Disability Syndrome. *American journal of human genetics*, *99*(3), 711–719.

Kuga, T., Kume, H., Adachi, J., Kawasaki, N., Shimizu, M., Hoshino, I., Matsubara, H., Saito, Y., Nakayama, Y., & Tomonaga, T. (2016). Casein kinase 1 is recruited to nuclear speckles by FAM83H and SON. *Scientific reports*, *6*, 34472.

Lenasi, T., & Barboric, M. (2013). Mutual relationships between transcription and pre-mRNA processing in the synthesis of mRNA. *Wiley interdisciplinary reviews. RNA*, *4*(2), 139–154.

Lesne, A., Baudement, M. O., Rebouissou, C., & Forné, T. (2019). Exploring Mammalian Genome within Phase-Separated Nuclear Bodies: Experimental Methods and Implications for Gene Expression. *Genes*, *10*(12), 1049.

Lester, E., Ooi, F. K., Bakkar, N., Ayers, J., Woerman, A. L., Wheeler, J., Bowser, R., Carlson, G. A., Prusiner, S. B., & Parker, R. (2021). Tau aggregates are RNA-protein assemblies that mislocalize multiple nuclear speckle components. *Neuron*, *109*(10), 1675–1691.e9.

Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R., & Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. *Cell*, *144*(1), 16–26.

Mao, Y. S., Zhang, B., & Spector, D. L. (2011). Biogenesis and function of nuclear bodies. *Trends in genetics : TIG*, *27*(8), 295–306.

Margaritis, T., & Holstege, F. C. (2008). Poised RNA polymerase II gives pause for thought. *Cell*, *133*(4), 581–584.

Marzahn, M. R., Marada, S., Lee, J., Nourse, A., Kenrick, S., Zhao, H., Ben-Nissan, G., Kolaitis, R. M., Peters, J. L., Pounds, S., Errington, W. J., Privé, G. G., Taylor, J. P., Sharon, M., Schuck, P., Ogden, S. K., & Mittag, T. (2016). Higher-order oligomerization promotes localization of SPOP to liquid nuclear speckles. *The EMBO journal*, *35*(12), 1254–1275.

Mattioni, T., Hume, C. R., Konigorski, S., Hayes, P., Osterweil, Z., & Lee, J. S. (1992). A cDNA clone for a novel nuclear protein with DNA binding activity. *Chromosoma*, *101*(10), 618–624.

Matveeva, E. A., Al-Tinawi, Q., Rouchka, E. C., & Fondufe-Mittendorf, Y. N. (2019). Coupling of PARP1-mediated chromatin structural changes to transcriptional RNA polymerase II elongation and cotranscriptional splicing. *Epigenetics & chromatin*, *12*(1), 15.

McArthur, E., & Capra, J. A. (2021). Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *American journal of human genetics*, *108*(2), 269–283.

McCord, R. P., Kaplan, N., & Giorgetti, L. (2020). Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Molecular cell*, *77*(4), 688–708.

Mintz, P. J., Patterson, S. D., Neuwald, A. F., Spahr, C. S., & Spector, D. L. (1999). Purification and biochemical characterization of interchromatin granule clusters. *The EMBO journal*, *18*(15), 4308–4320.

Misteli, T., Cáceres, J. F., Clement, J. Q., Krainer, A. R., Wilkinson, M. F., & Spector, D. L. (1998). Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *The Journal of cell biology*, *143*(2), 297–307.

Misteli, T., & Spector, D. L. (1999). RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Molecular cell*, *3*(6), 697–705.

Morgan G. T. (2002). Lampbrush chromosomes and associated bodies: new insights into principles of nuclear structure and function. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, *10*(3), 177–200.

Müller, W. G., Walker, D., Hager, G. L., & McNally, J. G. (2001). Large-scale chromatin decondensation and recondensation regulated by transcription from a natural promoter. *The Journal of cell biology*, *154*(1), 33–48.

Muse, G. W., Gilchrist, D. A., Nechaev, S., Shah, R., Parker, J. S., Grissom, S. F., Zeitlinger, J., & Adelman, K. (2007). RNA polymerase is poised for activation across the genome. *Nature genetics*, *39*(12), 1507–1511.

Nagashima, R., Hibino, K., Ashwin, S. S., Babokhov, M., Fujishiro, S., Imai, R., Nozaki, T., Tamura, S., Tani, T., Kimura, H., Shribak, M., Kanemaki, M. T., Sasai, M., & Maeshima, K. (2019). Single nucleosome imaging reveals loose genome chromatin networks via active RNA polymerase II. *The Journal of cell biology*, *218*(5), 1511–1530.

Nepon-Sixt, B. S., & Alexandrow, M. G. (2019). DNase I Chromatin Accessibility Analysis. *Bio-protocol*, *9*(23), e3444.

Nott, T. J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T. D., Bazett-Jones, D. P., Pawson, T., Forman-Kay, J. D., & Baldwin, A. J. (2015). Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular cell*, *57*(5), 936–947.

Pakravan, D., Orlando, G., Bercier, V., & Van Den Bosch, L. (2021). Role and therapeutic potential of liquid-liquid phase separation in amyotrophic lateral sclerosis. *Journal of molecular cell biology*, *13*(1), 15–28.

Pfaffl M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic acids research*, *29*(9), e45.

Phair RD, Misteli T. (2000). High mobility of proteins in the mammalian cell nucleus. *Nature*, *404*(6778), 604-609.

Quinodoz, S. A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J. M., Detmar, E., Lai, M. M., Shishkin, A. A., Bhat, P., Takei, Y., Trinh, V., Aznauryan, E., Russell, P., Cheng, C., Jovanovic, M., Chow, A., Cai, L., McDonel, P., Garber, M., & Guttman, M. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell*, *174*(3), 744–757.e24.

Rafalska-Metcalf, I. U., Powers, S. L., Joo, L. M., LeRoy, G., & Janicki, S. M. (2010). Single cell analysis of transcriptional activation dynamics. *PloS one*, *5*(4), e10272.

Raghuram, N., Strickfaden, H., McDonald, D., Williams, K., Fang, H., Mizzen, C., Hayes, J. J., Th'ng, J., & Hendzel, M. J. (2013). Pin1 promotes histone H1 dephosphorylation and stabilizes its binding to chromatin. *The Journal of cell biology*, *203*(1), 57–71.

Ren, J., Hathaway, N. A., Crabtree, G. R., & Muegge, K. (2018). Tethering of Lsh at the Oct4 locus promotes gene repression associated with epigenetic changes. *Epigenetics*, *13*(2), 173–181.

Sacco-Bubulya, P., & Spector, D. L. (2002). Disassembly of interchromatin granule clusters alters the coordination of transcription and pre-mRNA splicing. *The Journal of cell biology*, *156*(3), 425–436.

Saitoh, N., Spahr, C. S., Patterson, S. D., Bubulya, P., Neuwald, A. F., & Spector, D. L. (2004). Proteomic analysis of interchromatin granule clusters. *Molecular biology of the cell*, *15*(8), 3876–3890.

Sawyer, IA, Sturgill, D, Dundr, M. (2019). Membraneless nuclear organelles and the search for phases within phases. *WIREs RNA*. 10:e1514.

Schneider, N., Wieland, F. G., Kong, D., Fischer, A., Hörner, M., Timmer, J., Ye, H., & Weber, W. (2021). Liquid-liquid phase separation of light-inducible transcription factors increases transcription activation in mammalian cells and mice. *Science advances*, *7*(1), eabd3568.

Sharma, A. S. (2011). *Son is Essential for Nuclear Speckle Organization, Cell Cycle Progression and Pre-mRNA Splicing* [Doctoral dissertation, Wright State University]. OhioLINK Electronic Theses and Dissertations Center.

Sharma, A., Markey, M., Torres-Muñoz, K., Varia, S., Kadakia, M., Bubulya, A., & Bubulya, P. A. (2011). Son maintains accurate splicing for a subset of human pre-mRNAs. *Journal of cell science*, *124*(Pt 24), 4286–4298.

Sharma, A., Takata, H., Shibahara, K., Bubulya, A., & Bubulya, P. A. (2010). Son is essential for nuclear speckle organization and cell cycle progression. *Molecular biology of the cell*, *21*(4), 650–663.

Soboleva, T. A., Parker, B. J., Nekrasov, M., Hart-Smith, G., Tay, Y. J., Tng, W. Q., Wilkins, M., Ryan, D., & Tremethick, D. J. (2017). A new link between transcriptional initiation and pre-mRNA splicing: The RNA binding histone variant H2A.B. *PLoS genetics*, *13*(2), e1006633.

Spector, D. L., Fu, X. D., & Maniatis, T. (1991). Associations between distinct pre-mRNA splicing components and the cell nucleus. *The EMBO journal*, *10*(11), 3467–3481.

Spector, D. L., & Lamond, A. I. (2011). Nuclear speckles. *Cold Spring Harbor perspectives in biology*, *3*(2), a000646.

Sun, C. T., Lo, W. Y., Wang, I. H., Lo, Y. H., Shiou, S. R., Lai, C. K., & Ting, L. P. (2001). Transcription repression of human hepatitis B virus genes by negative regulatory element-binding protein/SON. *The Journal of biological chemistry*, *276*(26), 24059–24067.

Sun, Q., Hao, Q., & Prasanth, K. V. (2018). Nuclear Long Noncoding RNAs: Key Regulators of Gene Expression. *Trends in genetics : TIG*, *34*(2), 142–157.

Takata, H., Nishijima, H., Ogura, S., Sakaguchi, T., Bubulya, P. A., Mochizuki, T., & Shibahara, K. (2009). Proteome analysis of human nuclear insoluble fractions. *Genes to cells : devoted to molecular & cellular mechanisms*, *14*(8), 975–990.

Tena, J. J., & Santos-Pereira, J. M. (2021). Topologically Associating Domains and Regulatory Landscapes in Development, Evolution and Disease. *Frontiers in cell and developmental biology*, *9*, 702787.

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., Freier, S. M., Bennett, C. F., Sharma, A., Bubulya, P. A., Blencowe, B. J., Prasanth, S. G., & Prasanth, K. V. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Molecular cell*, *39*(6), 925–938.

Tripathi, V., Song, D. Y., Zong, X., Shevtsov, S. P., Hearn, S., Fu, X. D., Dundr, M., & Prasanth, K. V. (2012). SRSF1 regulates the assembly of pre-mRNA processing factors in nuclear speckles. *Molecular biology of the cell*, *23*(18), 3694–3706.

Uversky V. N. (2017). Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Current opinion in structural biology*, *44*, 18–30.

Vaňková Hausnerová, V., & Lanctôt, C. (2017). Chromatin decondensation is accompanied by a transient increase in transcriptional output. *Biology of the cell*, *109*(1), 65–79.

Varia, S., Potabathula, D., Deng, Z., Bubulya, A., & Bubulya, P. A. (2013). Btf and TRAP150 have distinct roles in regulating subcellular mRNA distribution. *Nucleus (Austin, Tex.)*, *4*(3), 229–240.

Wagner, R. (1835). Einige bemerkungen und fragen über das keimbläschen (vesicular germinativa). *Müller's Archiv Anat Physiol Wissenschaft Med*, *268*, 373-7.

Wang, H., Han, M., & Qi, L. S. (2021). Engineering 3D genome organization. *Nature reviews. Genetics*, *22*(6), 343–360.

Wang, R., & You, J. (2015). Mechanistic analysis of the role of bromodomain-containing protein 4 (BRD4) in BRD4-NUT oncoprotein-induced transcriptional activation. *The Journal of biological chemistry*, *290*(5), 2744–2758.

Wegmann, S., Eftekharzadeh, B., Tepper, K., Zoltowska, K. M., Bennett, R. E., Dujardin, S., Laskowski, P. R., MacKenzie, D., Kamath, T., Commins, C., Vanderburg, C.,

Roe, A. D., Fan, Z., Molliex, A. M., Hernandez-Vega, A., Muller, D., Hyman, A. A., Mandelkow, E., Taylor, J. P., & Hyman, B. T. (2018). Tau protein liquid-liquid phase separation can initiate tau aggregation. *The EMBO journal*, *37*(7), e98049.

Wei, L., Nakajima, S., Böhm, S., Bernstein, K. A., Shen, Z., Tsang, M., Levine, A. S., & Lan, L. (2015). DNA damage during the G0/G1 phase triggers RNA-templated, Cockayne syndrome B-dependent homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(27), E3495–E3504.

Wynn, S. L., Fisher, R. A., Pagel, C., Price, M., Liu, Q. Y., Khan, I. M., Zammit, P., Dadrah, K., Mazrani, W., Kessling, A., Lee, J. S., & Buluwela, L. (2000). Organization and conservation of the GART/SON/DONSON locus in mouse and human genomes. *Genomics*, *68*(1), 57–62.

Yang, L., Lin, C., Liu, W., Zhang, J., Ohgi, K. A., Grinstein, J. D., Dorrestein, P. C., & Rosenfeld, M. G. (2011). ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell*, *147*(4), 773–788.

Youmans, D. T., Schmidt, J. C., & Cech, T. R. (2018). Live-cell imaging reveals the dynamics of PRC2 and recruitment to chromatin by SUZ12-associated subunits. *Genes & development*, *32*(11-12), 794–805.

Zbinden, A., Pérez-Berlanga, M., De Rossi, P., & Polymenidou, M. (2020). Phase Separation and Neurodegenerative Diseases: A Disturbance in the Force. *Developmental cell*, *55*(1), 45–68.

Zidovska A. (2020). The self-stirred genome: large-scale chromatin dynamics, its biophysical origins and implications. *Current opinion in genetics & development*, *61*, 83–90.

Zidovska, A., Weitz, D. A., & Mitchison, T. J. (2013). Micron-scale coherence in interphase chromatin dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(39), 15555–15560.

Zhao, R., Nakamura, T., Fu, Y., Lazar, Z., & Spector, D. L. (2011). Gene bookmarking accelerates the kinetics of post-mitotic transcriptional re-activation. *Nature cell biology*, *13*(11), 1295–1304.

ZTTK SON-Shine Foundation. (n.d.). Retrieved from https://zttksonshinefoundation.org/

Zuin, J., Dixon, J. R., van der Reijden, M. I., Ye, Z., Kolovos, P., Brouwer, R. W., van de Corput, M. P., van de Werken, H. J., Knoch, T. A., van IJcken, W. F., Grosveld, F. G., Ren, B., & Wendt, K. S. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(3), 996–1001.

APPENDIX A

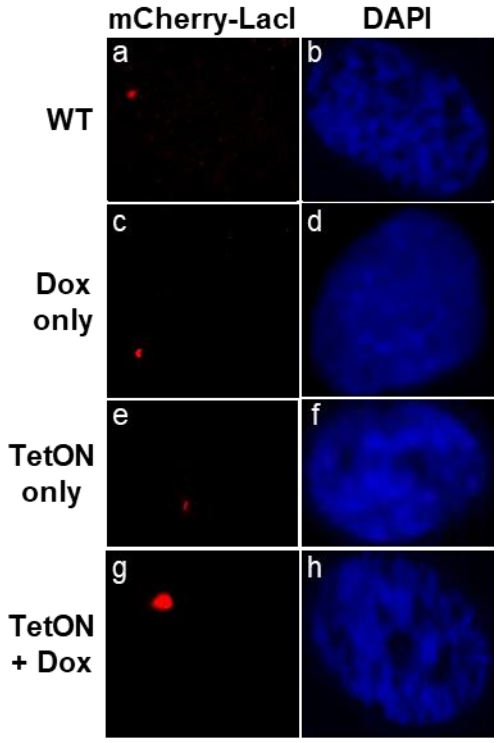
U2OS 2-6-3 REPORTER CELL LINE CHARACTERIZATION

U2OS 2-6-3 cells with stably incorporated mCherry-LacI (mU2OS) were used to visualize the reporter gene locus since using this cell line eliminates the need for a LacI-fluorescent protein transfection step. mU2OS cells were left as wild-type uninduced and untransfected (no pTetON – dox; Figure A1; panels a-b), untransfected with doxycycline added (no pTetON + dox; Figure A1; panels c-d), transfected with pTetON without doxycycline (pTetON – dox; Figure A1; panels e-f), or transfected with pTetON and doxycline added (pTetON + dox; Figure A1; panels g-h). As expected, only the "pTetON + dox" condition showed visible mCherry-LacI labelled locus decondensation (Figure A1).

As first described in Janicki et al., 2004, accumulation of reporter transcript is first visible by microscopy within 7.5 min post-dox. A YFP-tagged MS2 binding protein (YFP-MS2BP) binds to the transcribed MS2 stem loop translational operators allowing visualization of reporter gene transcription by microscopy. Within 30 min after reporter gene activation, noticeable changes in higher order chromatin structure can be detected, with maximal chromatin decondensation approximately 120 min post-dox. YFP-MS2BP intensity increases at the transcription site until 150 min post-dox, when a decrease in intensity is detected (Janicki et al., 2004).

<u>Figure A1:</u> mCherry-LacI labelled U2OS 2-6-3 loci show visible decondensation with pTetON only in the presence of doxycycline.

U2OS 2-6-3 cells stably expressing mCherry-LacI were either left untransfected and uninduced (no pTetON – dox; a-b), untransfected with dox (no pTetON + dox; c-d), or transfected with TetON and uninduced (pTetON - dox; e-f) or induced with dox (pTetON + dox; g-h). At 2.5 h after dox addition, the cells were processed for immunofluorescence. DNA was stained with DAPI.



I wanted to replicate these findings to determine the reporter mRNA kinetics in mU2OS cells. mU2OS cells were transfected with YFP-MS2BP and TetON, followed by increasing timepoints of dox treatment (Figure A2). “NoTetON – dox” cells (Figure A2; panels a-d) showed diffuse YFP-MS2BP nuclear localization without accumulation at the mU2OS locus. When YFP-MS2BP is transfected into the mU2OS cells, it shows a diffuse nuclear localization because reporter RNA containing MS2 repeats is not present. At 5 min post-dox (Figure A2; panels e-h) and 10 min post-dox (Figure A2; panels i-l) a small amount of YFP-MS2BP has accumulated at the still-condensed mU2OS locus, although most YFP-MS2BP is still diffused throughout the nucleus. By 30 min post-dox (Figure A2; panels m-p), the mU2OS locus is visibly surrounded by YFP-MS2BP, with YFP-MS2BP accumulating at the active transcription site. YFP-MS2BP continues to accumulate at the mU2OS locus at timepoints 60 min dox, 90 min dox, and 120 min dox (Figure A2; panels q-t; panels u-x; panels y-bb).

Cells on each slide were imaged individually and exposure time varies between conditions. To prevent oversaturation, exposure was especially decreased beginning at the 30 min post-dox timepoint, causing the nucleus to appear much darker than the previous timepoints. Although YFP-MS2BP does begin to be more robustly depleted from the surrounding nucleus for locus recruitment at this time, the images presented are shown with slightly darker nuclei than would be visible by eye via microscopy. Similarly, the mU2OS locus appears oversaturated in the “no TetON – dox,” “5 min dox,” and 10 min dox” images shown.

Reporter transcript production was measured by qRT-PCR to rule out leaky transcription by quantitative analysis of reporter gene expression. Primer pairs in the

reporter intron, exon, intron-exon junction, and CFP open reading frame were designed by undergraduate researcher Jacob Ward. All primer sets showed a robust increase in reporter transcript in the active "TetON + dox" conditions when normalized to GAPDH and the inactive "no TetON – dox" condition, with the largest increase using the primer set in the CFP open reading frame (Figure A3a). A melt curve for each of the primer sets is displayed in Figure A3b, and a single peak (as shown) signals the primers each amplify a single PCR product. Only "TetON + dox" condition showed a robust fold-change of 14.328 in reporter transcript normalized to GAPDH and mock transfected controls (Figure A3c), indicating that reporter transcription activity was detectable only at induced loci.

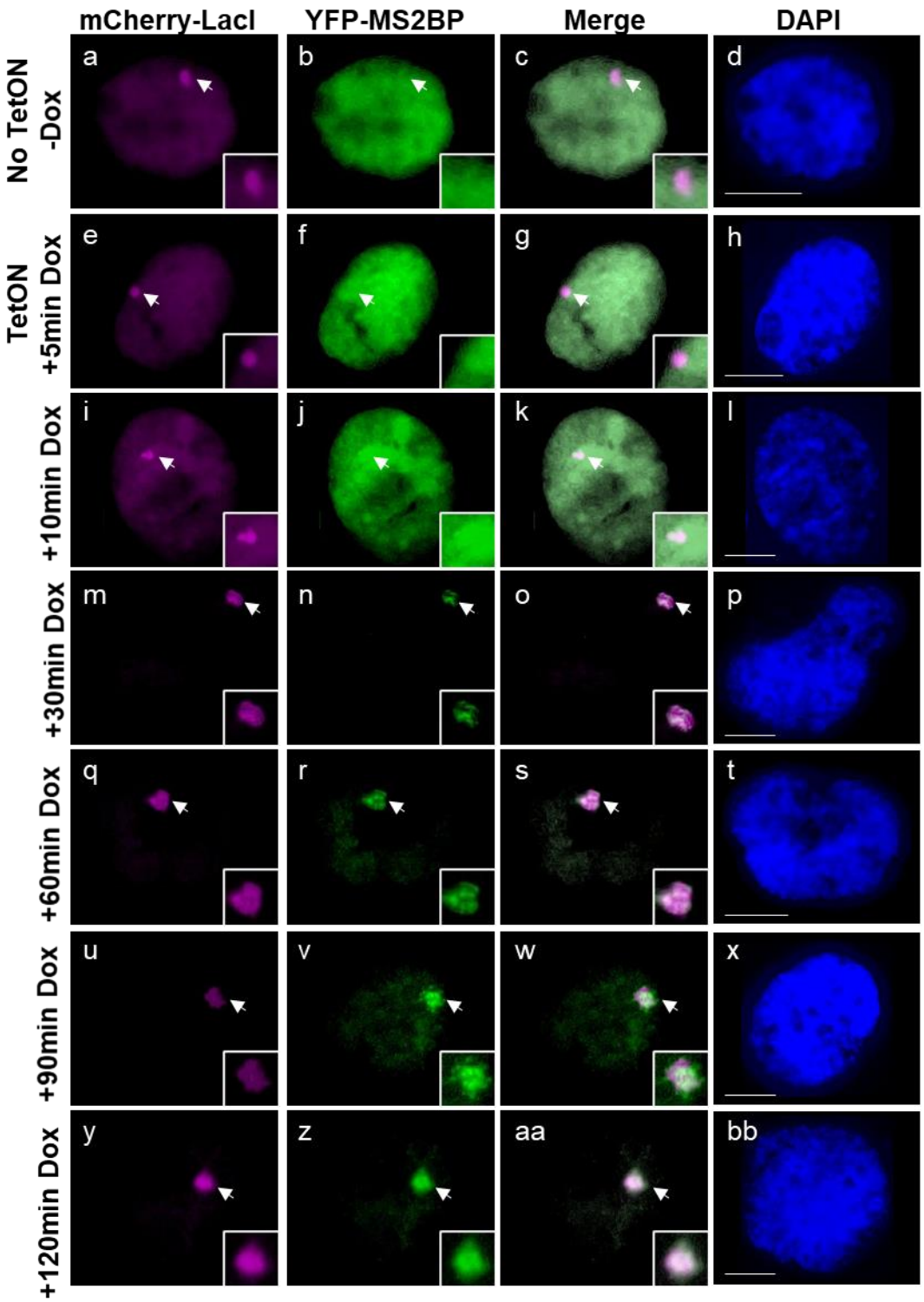
Microscopy analysis of the inactive and active mU2OS locus indicates the inactive locus is heterochromatic. Janicki et al., 2004 identified the heterochromatic epigenetic marker H3K9me3 and HP1α, which binds to H3K9me3 (Bannister et al., 2001; Jacobs et al., 2001) at the inactive mU2OS locus. Indeed, H3K9me3 was colocalized with the inactive reporter locus in mU2OS cells (Figure A4a), indicating the locus is heterochromatic in its OFF state.

When transcription is activated with TetON in the presence of doxycycline, epigenetic markers including H4K5ac, H4K12ac and H3K9ac are present at the decondensed transcription site promoter, along with their associated histone acetyltransferases GCN5, PCAF, and p300 (Rafalska-Metcalf et al., 2010). Pre-mRNA processing factors Btf and TRAP150, SRSF1, and Magoh have been shown to associate with the active reporter locus (Varia et al., 2013). In mU2OS cells, Btf and TRAP150 also colocalize with the active transcription site (Figure A4b). There are approximately

200 copies of the reporter construct integrated at the U2OS 2-6-3 locus, and there is also physical separation between the Lac operator repeats that are labelled with mCherry-LacI and the downstream transcription unit that recruits pre-mRNA processing factors. So Btf and TRAP150 staining of active cells shows recruitment to the area near mCherry-LacI, but is often visually distinct from mCherry-LacI foci.

<u>Figure A2:</u> Reporter transcript can be detected within 30 minutes after transcription induction.

Timecourse using U2OS 2-6-3 cells with stably incorporated mCherry-LacI. Cells in 7 different wells were simultaneously electroporated and first dox treatment (150 min) added 3 hours post-transfection. Dox was added at decreasing timepoints until all cells were harvested 5 hours post-transfection and cells processed for immunofluorescence. DNA was stained with DAPI. Scale bar = 5 μm.


mCherry-LacI
YFP-MS2BP
Merge
DAPI
No TetON -Dox
TetON +5min Dox
+10min Dox
+30min Dox
+60min Dox
+90min Dox
+120min Dox
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z
aa
bb

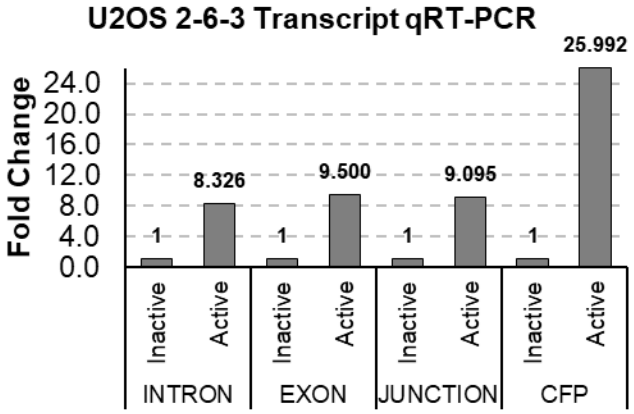
<u>Figure A3:</u> Reporter transcript is upregulated in transcriptionally active TetON + dox conditions.

A. qRT-PCR analysis shows relative fold change in U2OS 2-6-3 reporter transcript abundance when normalized to GAPDH and "no TetON – dox" controls. Primer sets were designed in the intron, exon, intron-exon junction, and CFP open reading frame regions of the U2OS 2-6-3 gene. Relative fold change was calculated by ΔΔCt method from the average Ct of 3 technical replicates from n = 1 biological replicate.

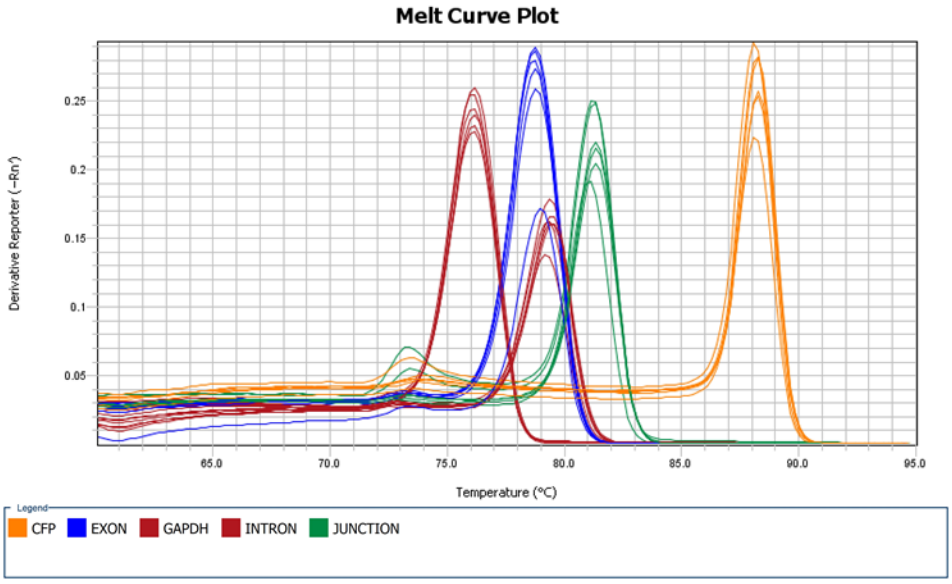
B. Melt curve plot from the QuantStudio 7 flex software displayed as negative derivative reporter ($-R_n'$) vs temperature for all PCR primer sets tested. The derivative reporter represents the rate of change in fluorescence as a function of temperature. Peaks in the melt curve indicate the melting temperature (T_m) of the primer set. T_m is the temperature at which 50% of the DNA is double stranded and 50% is single stranded. Multiple T_m peaks would indicate nonspecific amplification such as the presence of primer dimers or multiple PCR products (Applied Biosystems User Guide)

C. qRT-PCR analysis shows relative fold change in U2OS 2-6-3 transcript abundance when nomalized to GAPDH and mock transfected control. mU2OS cells were either left untransfected and uninduced (WT), transfected with carrier DNA and uninduced (mock), induced with dox (dox only), transfected with TetON and uninduced (TetON only) or induced with dox (TetON+dox). At 5.5 hrs post-transfection, total RNA was extracted and two-step qRT-PCR performed. Relative fold change was calculated by ΔΔCt method from the average Ct of 3 technical replicates from n = 1 biological replicate.

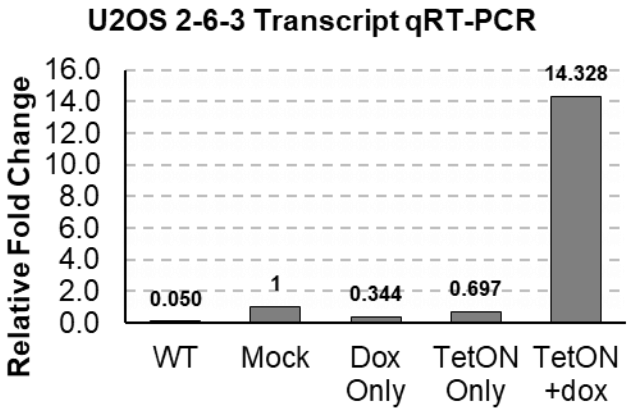
A


U2OS 2-6-3 Transcript qRT-PCR
Fold Change
24.0
20.0
16.0
12.0
8.0
4.0
0.0
1
8.326
1
9.500
1
9.095
1
25.992
Inactive
Active
INTRON
EXON
JUNCTION
CFP

B


Melt Curve Plot
Derivative Reporter (-Rn')
0.25
0.2
0.15
0.1
0.05
65.0
70.0
75.0
80.0
85.0
90.0
95.0
Temperature (°C)
Legend
CFP
EXON
GAPDH
INTRON
JUNCTION

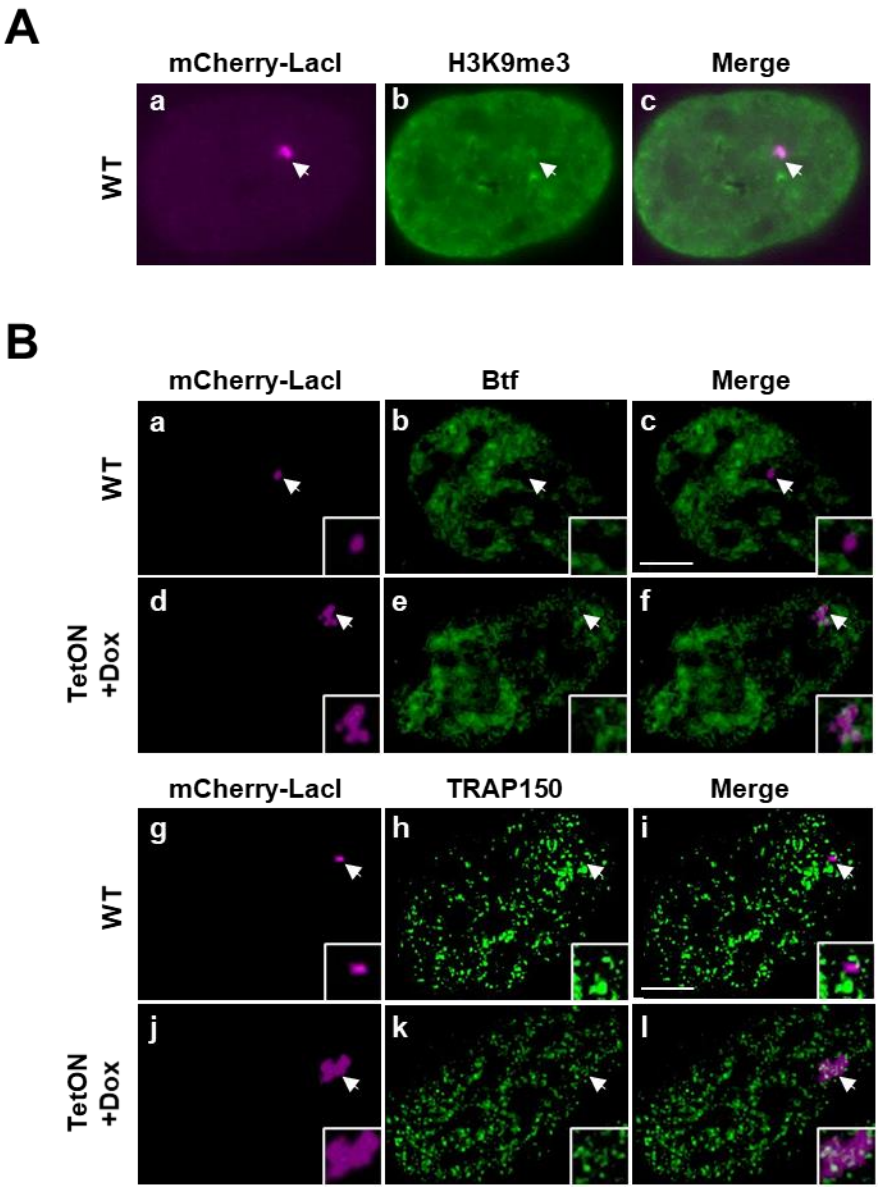
C


U2OS 2-6-3 Transcript qRT-PCR
Relative Fold Change
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
0.050
1
0.344
0.697
14.328
WT
Mock
Dox Only
TetON Only
TetON +dox

<u>Figure A4</u>: The inactive mU2OS locus is heterochromatic and recruits pre-mRNA processing factors upon transcription activation.

A. H3K9me3 colocalizes with the closed mU2OS locus (16/20 cells scored). Wild type cells expressing mCherry-lacI were processed for immunofluorescence and colocalization with H3K9me3.

B. Pre-mRNA processing factors localize to the transcriptionally active mU2OS locus. Cells were left untransfected and uninduced (WT) or transfected with TetON + 2.5 hrs dox. Post-transfection 5.5 hrs, cells were processed for immunofluorescence and colocalization with Btf or TRAP150. Btf was not colocalized with WT mCherry-LacI in 28/30 cells scored (panels a-c), and colocalized with the transcriptionally active locus in 25/30 cells scored (panels d-f). TRAP150 was not colocalized with WT mCherry-LacI in 27/30 cells scored (panels g-i), and colocalized with the transcriptionally active locus in 26/30 cells scored (panels j-l).


A
mCherry-LacI
H3K9me3
Merge
WT
a
b
c
B
mCherry-LacI
Btf
Merge
WT
TetON +Dox
a
b
c
d
e
f
mCherry-LacI
TRAP150
Merge
WT
TetON +Dox
g
h
i
j
k
l

APPENDIX B

```
clc
clear all %#ok<*CLALL>
prefix = {'a'};    %Define the letters that represent the different experimental portions for file naming
indices = [101];  %How many of each letters are there, put them in order matchin g prefix
PixRedFactor = 4; %PR = 4
SobelThresh = 0.03; %ST = 0.03
for R = 1:length(prefix)
for RR = indices(R)
name1 = join([num2str(RR),prefix(R)],''); %Joins the numberical value and letter for making a complete filename for reading
name = join([name1,'.tif'],'');
names(RR,R) = join([num2str(RR),prefix(R)],'');   %Places all names into a vector for output file use
filenames = dir(char(name)); %Reads the image file
PrintI2 = 0;
PrintI3 = 0;
PrintI4 = 0;
PrintI5 = 0;
```

```
PrintI6 = 0;

PrintI7 = 0;

PrintA1 = 0;

PrintI8 = 0;

PrintI9 = 0;

PrintI10 = 0;

PrintI11 = 0;

PrintI12 = 0;

PrintI13 = 0;

PrintIndex = [PrintI2   PrintI3   PrintI4   PrintI5   PrintI6   PrintI7   PrintA1   PrintI8

PrintI9   PrintI10   PrintI11   PrintI12   PrintI13]';

s = numel(filenames);

 [PrintNameList] = GeneratingPrintName(PrintIndex,s);

Arealist = zeros(s,1);

edgecountlist = zeros(s,1);

edgedenlist = zeros(s,1);

for q=1:numel(filenames)

%Load image

I = imread(filenames(q).name);

I = imresize(I,[512   512]);

%Acquire threshold value for I

 [T] = ThreshMode(I);

%Image average smoothening by (i)th times
```

```
I2 = I;
for i = 1:6
h = fspecial('average');
I2 = imfilter(I2,h);
end
%Thereshold application to I2
[I3] = ApplyThresh(I2,T);
I3 = logical(I3);
%Hole-filling algorithm
I4 = imfill(I3,'holes');
%Extract the nucleus from the original image to a black background
[I5] = ExtractImage(I,I4);
I5 = uint8(I5);
%Intensity redistribution for I5
A = max(max(I5));
B = double(I5);
C = double(A);
I5 = (B/C)*255;
I5 = uint8(I5);
%Image reduction by a factor of 4 (1/4 = 0.25)
PixRed = 1/PixRedFactor;
I6 = imresize(I5,PixRed);
%Intensity redistribution for I6
```

```
A = max(max(I6));

B = double(I6);

C = double(A);

I6 = (B/C)*255;

I6 = uint8(I6);

%SOBEL edge detection application

I7 = edge(I6,'sobel',SobelThresh);

A1 = uint8(I7);

A1 = A1*255;

%Acquire threshold value for I6

clear T

[T] = ThreshMode(I6);

%Threshold application to I6

[I8] = ApplyThresh(I6,T);

I8 = logical(I8);

%Hole-filling algorithm

I9 = imfill(I8,'holes');

%Perimeter subtraction by (n)th times

I10 = I9;

n = 2;

for i = 1:n

I11 = bwperim(I10);

I10 = I10-I11;
```

```
I10 = logical(I10);
end
%Extract the SOBEL edge inside the nucleus into a black background
 [I12] = ExtractImage(I7,I10);
I12 = logical(I12);
I13 = I12+I11;
I13 = uint8(I13);
I13 = I13*255;
%Nucleus area
[row,column,int] = find(I10>0);
S = length(row);
Area = S;
Arealist(q,1) = Area;
%Edge count
edgecount = sum(sum(I12));
edgecountlist(q,1) = edgecount;
%Edge density (i.e. chromatin condensation parameter)
edgeden = (edgecount/Area)*100;
edgedenlist(q,1) = edgeden;
I12 = uint8(I12);
I12 = I12*255;
 [PrintList] = GeneratingPrint(PrintIndex,q,PrintNameList,I2,I3,I4,...
I5,I6,I7,A1,I8,I9,I10,I11,I12,I13);
```

```
Arealist1(RR,R) = Arealist;
edgecountlist1(RR,R) = edgecountlist; %#ok<*SAGROW>
edgedenlist1(RR,R) = edgedenlist;
figure
subplot(2,3,1)
imshow(I)
title('Original Image')
subplot(2,3,2)
imshow(I2)
title('Smoothed Image')
subplot(2,3,3)
imshow(I5)
title('Intensity Redistribution Image')
subplot(2,3,4)
imshow(I6)
title('Complete Background Removed Image')
subplot(2,3,5)
imshow(I12)
title('Sobel edge detection of interior of cell')
subplot(2,3,6)
imshow(I13)
title('Total cell perimeter with sobel detection of interior')
end
```

```
%%
% figure
% subplot(2,3,1)
% imshow(I)
% title('Original Image')
% subplot(2,3,2)
% imshow(I2)
% title('Smoothed Image')
% subplot(2,3,3)
% imshow(I5)
% title('Intensity Redistribution Image')
% subplot(2,3,4)
% imshow(I6)
% title('Complete Background Removed Image')
% subplot(2,3,5)
% imshow(I12)
% title('Sobel edge detection of interior of cell')
% subplot(2,3,6)
% imshow(I13)
% title('Total cell perimeter with sobel detection of interior')
```