THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

THE EFFECT OF SUBUNIT DELETIONS ON SAGA COMPLEX BINDING IN *SACHAROMYCES CEREVISIAE*

MATTHEW DRIBAN
SPRING 2019

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Biochemistry and Molecular Biology
with honors in Biochemistry and Molecular Biology

Reviewed and approved* by the following:

B. Franklin Pugh
Willaman Chair in Molecular Biology and Professor of Biochemistry and Molecular Biology
Thesis Supervisor

Craig Cameron
Professor of Biochemistry and Molecular Biology
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

The SAGA complex is a multi-functional protein complex primarily involved in transcriptional regulation and histone modification, particularly under stress conditions. Notably the SAGA complex has also been linked to development of certain neurodegenerative diseases and cancers, making characterization of SAGA in peril a priority. Here, to better characterize SAGA's ability to regulate gene expression under various conditions of stress, I analyzed SAGA's binding characteristics under two conditions in combination: subunit knockouts and heat shock. These analyses resulted in a clearer picture of SAGA's gene regulatory role under different conditions of stress, lending insights into several critical disease phenotypes. In particular, the importance of the subunits Gcn5, Spt20, and Ahc1 in recruiting transcription pre-initiation complex machinery were elucidated. Together, these results yield better characterization of SAGA's mechanistic response to stress, especially in shedding light on Gcn5 and Spt20's critical importance in TBP recruitment.

TABLE OF CONTENTS

LIST OF FIGURES iii

ACKNOWLEDGEMENTS v

Chapter 1. Introduction 1

Chapter 2. Materials and Methods 7

Chapter 3. Results 17

Binding Datasets are Consistent Across Biological Replicates 17
Stress-Induced PIC Assembly is Most Strongly Upregulated in SAGA-Mediated Genes 18
Gcn5 and Spt20 Deletions Impair TBP Recruitment to the TSS 25
Ahc1 Deletions Do Not Affect PIC Assembly 26

Chapter 4. Discussion 28

Appendix A Reagents 32

Appendix B Reaction Specifications 39

BIBLIOGRAPHY 45

ACADEMIC VITA 48

LIST OF FIGURES

Figure 1. 2D Architecture of the SAGA Complex....2

Figure 2. Generation of KanMx Knockouts via a PCR Homology Approach....4

Figure 3. ChIP-exo Workflow....6

Figure 4. Composite Plots of Control TBP Biological Replicates....17

Figure 5. Genome Wide and SAGA-Mediated Recruitment of TBP and Pol II by the SAGA Complex Under Normal and Heat Shock Growth Conditions....24

Figure 6. TBP and Pol II Recruitment by Gcn5 and Spt20-Deletion Strains....26

Figure 7. SAGA Response to an Ahc1-Deficient ADA Complex....27

Figure 8. Gcn5 and Spt20-Mediated Recruitment of TBP....31

ACKNOWLEDGEMENTS

Thank you to all members of the Pugh Lab for their assistance throughout my thesis project. Special thanks to Dr. Chitvan Mittal and Dr. Alain Bataille for their patience and expertise in preparing me technically and scientifically for my project. A further thank you to Jordan John for his willingness to lend a hand and an ear when needed. Finally, I am most grateful for Dr. Frank Pugh's guidance in my project and for my personal and academic growth with his group.

Chapter 1. Introduction

Transcription is gene expression's first step, generating a single-stranded RNA molecule from a DNA template. Its initiation and progression are mediated by a diverse set of proteins and protein complexes. One such protein complex is the Spt-Ada-Gcn5 acetyltransferase (SAGA) complex, which has several regulatory roles and is highly conserved across eukaryotic species[1]. The ADA complex is another protein complex with similar regulatory roles and some shared subunits to the SAGA complex[2]. Understanding the function of the SAGA complex is of critical importance because of its diverse set of functionalities in regulating gene expression, with histone acetyl transferase (HAT), deubiquitinase (DUB), TATA-binding protein (TBP) binding, nucleosome binding, and activator binding roles[3]. Across the entire yeast genome, SAGA helps regulate about 10% of genes, the majority of which become more highly transcribed under stress conditions[4]. Shedding light on how the SAGA complex and SAGA-mediated genes deal with stress may better characterize gene regulation and stress response in eukaryotic organisms.

Here, I use *Saccharomyces cerevisiae*, budding yeast, as a model system to study SAGA's role in gene regulation. Yeast provides a relatively efficient avenue to study eukaryotic gene regulation, many features of which are consistent across species. In particular, simulating the stress conditions that this thesis aims to study in the SAGA complex is much easier using the tools available in yeast genetics compared to mammalian genetics. Ease of use does not sacrifice relevance to the human genome however, as yeast genetics have been consistently demonstrated to have applicability to mammalian genome. 23% of yeast genes have a human homolog and many more cellular processes have significant crossover with a human process[5].

Yeast contains one SAGA complex and humans contain four SAGA-like complexes, with overlapping roles. It is likely that these four SAGA-like complexes evolved from one complex via duplication and mutation[1]. Yeast's SAGA complex is 1.8 MDa and has 18-20 subunits divided into specific structural modules (Figure 1)[3]. In yeast these modules include the HAT, DUB, core structural, TBP binding, and transcription factor (TF) binding modules, while the mammalian SAGA complex also includes a splicing module[6]. Importantly, the majority of these subunits have a functional equivalent across eukaryotic species meaning SAGA's regulatory roles remain fairly consistent.

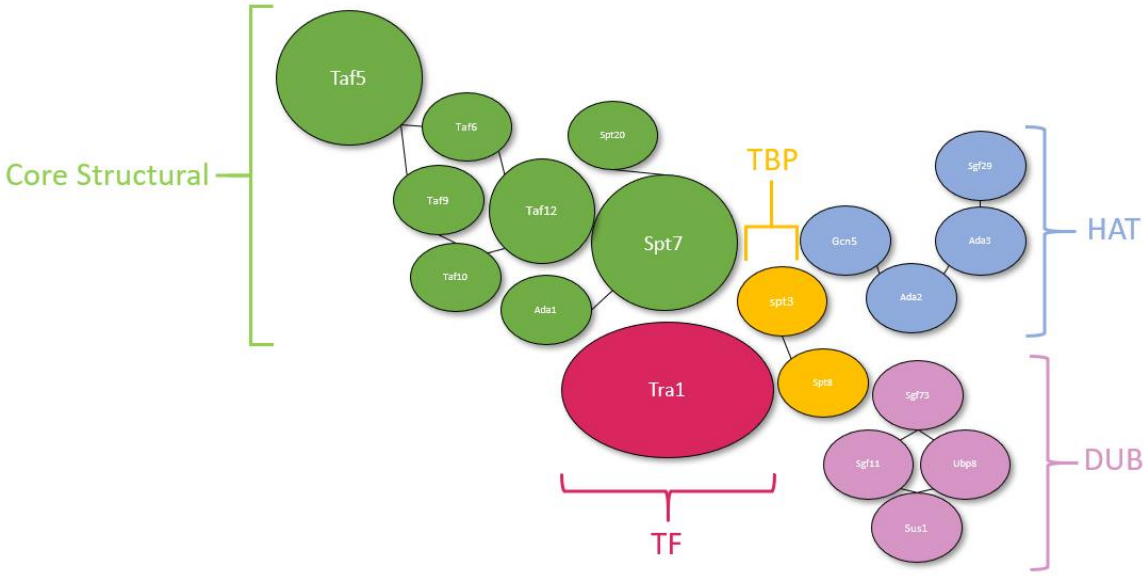


Figure 1. 2D Architecture of the SAGA Complex

The SAGA complex consists of 18-20 subunits divided into 5 modules: core structural, transcription factor binding (TF), HAT, DUB, and TBP binding (TBP)[7].

SAGA's roles in transcription pre-initiation complex (PIC) assembly and RNA polymerase II (Pol II) recruitment have been previously studied, but remain incompletely characterized[8]. PIC assembly is the process by which TBP and several general transcription

factors (GTFs) come together to recruit Pol II and begin transcription. Besides having a TBP-binding module, SAGA has also been reported to be a general cofactor for Pol II transcription and to help initiate PIC assembly with its HAT activity, acetylating lysine residues on histone H3[8, 9]. Better understanding of these behaviors in the early stages of transcription, particularly under abnormal cell conditions, may yield new insights into SAGA's regulatory role.

Understanding SAGA's ability to function under conditions of stress is of critical importance because of its connotations in human disease development. Previous research has linked Gcn5, the HAT subunit in SAGA, with neural functionality. Gcn5 depletion significantly accelerated cerebellar and retinal degeneration in mice with spinal cerebellar ataxia type 7 (SCA7), a neurodegenerative disorder. Another subunit, Sgf29, which recognizes methylated histones and recruits the SAGA complex to site has been linked to the Myc family of cancers, promoting c-Myc-mediated malignancies[10]. A third, Usp22, part of the DUB module of the mammalian SAGA complex, has been correlated with poor prognosis in several types of cancers[11]. As a unicellular organism, yeast may not provide a direct avenue into oncogenic studies, but its ability or inability to adapt to deficiencies in the SAGA complex may yield insights into disease development nevertheless, since many core regulatory pathways are the same in humans and yeast.

To assess the effect of damage to the SAGA complex on its function, I use subunit deletions and heat shock stress in collaboration. Each selected subunit deletion was chosen for the purpose of removing a core SAGA function, forcing the cells to reprogram in response to the knockout. This adaptation process is most acute under heat shock stress, where SAGA is most active in promoter binding. First, subunit knockouts were generated for selected subunits of the SAGA complex in a wild type background (Figure 2). However, given that SAGA is

predominantly responsive to environmental stress conditions, a second stressor was deemed necessary to accurately and robustly delve into potential disease phenotypes for the complex. Heat shock was selected for this role[12, 13]. Comparison of binding behaviors under normal growth conditions to heat shock was intended to reflect SAGA's reported increase in transcriptional activity under environmental stress, predicting a stronger signal under heat shock stress. Pairing this amplified signal with subunit knockouts was intended to assess SAGA's sustained ability to assemble the transcription machinery in peril.

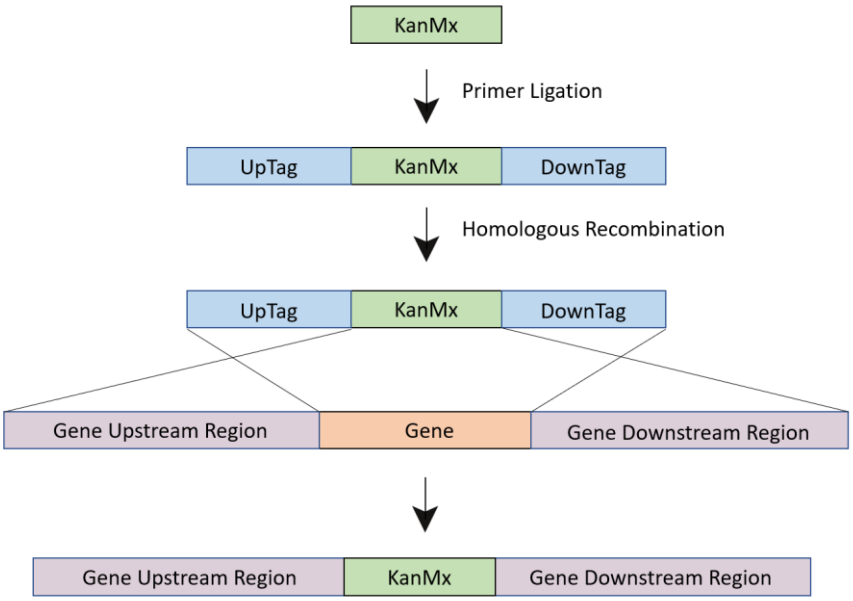


Figure 2. Generation of KanMx Knockouts via a PCR Homology Approach

Gene knockouts were generated via a PCR-based gene deletion strategy, adapted from the Stanford Saccharomyces Genome Deletion Project (http://www-sequence.stanford.edu/group/yeast_deletion_project/deletions3.html). Using this approach, regions of upstream and downstream homology to the gene of interest recombine with flanking regions to the KanMx deletion marker, creating a knockout strain selectable via exposure to kanamycin.

The subunits selected for knockout are as follows: Gcn5, Spt20, Sgf29, Sgf73, and Ahc1. Gcn5, as the HAT subunit, is important to SAGA's ability to acetylate chromatin and recruit the PIC complex. Deletion is likely to decrease levels of both events. Spt20 maintains structural integrity in the complex[14]. Deletion of Spt20 may result in complete dissolution of the complex or, if other structural integrity subunits are able to compensate, may have minimal effect on SAGA's ability to bind DNA. Sgf29 is part of the HAT module and binds to Ada3, a subunit that mediates Gcn5 activity[15]. Sgf73 is part of the DUB module, absence of which may be less apparent in DNA binding but more important in loss of deubiquitination in the cell[16]. Finally, Ahc1 is a critical structural integrity subunit in SAGA's sister complex, the ADA complex, which also has HAT functionality[2]. Disappearance of Ahc1 may show a compensatory effect in the SAGA complex, if ADA is unable to bind and perform its cellular duty. These five subunits, under stressed and normal growth conditions, present a diverse look into how mutations in different SAGA modules affect its ability to bind to DNA and regulate transcription.

Since DNA binding was selected as the proxy to analyze SAGA's mutant phenotype, chromatin immunoprecipitation with an exonuclease digestion (ChIP-exo) was used as the primary technique in this study, after sample preparation (Figure 2). ChIP in general is the most popular tool to assess protein binding locations genomewide[17]. It works by processing cellular material with proteins crosslinked to DNA. After a series of washes and ligations to prime material for sequencing analysis, the proteins are eluted, and the DNA is purified and sequenced. ChIP-exo introduces a critical step absent in other ChIP protocols, with a lambda nuclease enzymatic 5'-3' digestion of non-crosslinked DNA prior to elution. This digests unbound DNA to approximately 6 base pairs (bp) away from the protein binding site, resulting in peak pairs to yield ultra-precise resolution of the protein's binding site[18]. Data on these binding sites,

including occupancy levels and sequence-specific information, can be used to assess the ability of SAGA to bind to DNA and overcome complex mutations. Insights into these binding events under subunit deletions, with and without heat shock stress, better illustrates the role SAGA has in PIC assembly, Pol II recruitment, and general transcriptional regulation.

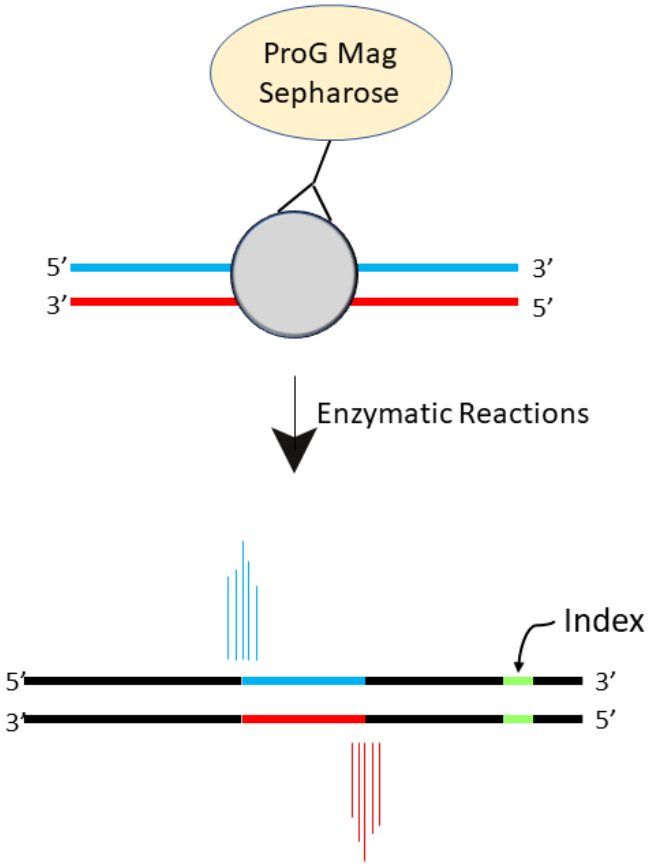


Figure 3. ChIP-exo Workflow

ChIP-exo 5.0 processes a crosslinked protein-DNA complex into a high-resolution binding sequence using magnetic beads. After washing the sample, the complex is enzymatically modified through multiple ligations, reverse crosslinking, PCR amplification, and a 5'→3' exonuclease digestion. This processing results in peak pair pileups at the 5' end of the protein on both DNA strands. Bioinformatic analysis of the peak pairs yields highly precise information about the protein binding location between the two peak pairs[19].

Chapter 2. Materials and Methods

Kanamycin Cassette Preparation

Transformation

A tube of New England Buffer (NEB) 10-beta Competent *E. coli* cells (NEB, cat#C3019H) was thawed on ice for 10 minutes. After thawing completely, 1 µl of the pFA6a-FRB-FRB-GFP-kanMX6 (p30580) plasmid was added to the cell tube, flicking gently to mix the cells and DNA. The mixture was then incubated on ice for 30 minutes, after which point it was heat shocked at 42℃ for 30 seconds. Then the mixture was cooled on ice for 5 minutes. Next, 950 µl of S.O.C. medium was added to the mixture and shaken in an incubator at 37℃ and 250 rpm to allow ample time for cell recovery and plasmid integration. While shaking, 2 LB-Ampicillin (LB-Amp) and 2 LB-Kanamycin (LB-Kan) selection plates were warmed to 37℃ (See Appendix A for reagent specifications). After 60 minutes, the cells were removed from the incubator. 20 µl of cells were spread on the first LB-Amp plate and 100 µl on the second LB-Amp plate, repeating for the 2 LB-Kan plates. These plates were then incubated at 37℃ overnight.

Purification

After overnight incubation, 4 colonies were colonies were selected per condition, LB-Amp or LB-Kan, and inoculated in 5 ml LB cultures. These cultures were grown overnight at 37℃ and 250 rpm. After overnight growth, each sample was harvested by centrifugation at room temperature (RT) for 7 minutes at 3,000 rpm. The plasmid DNA from each sample was then purified using the QIAprep Spin Miniprep Kit protocol.

Confirmation

To determine whether the purified plasmid was the desired p30580 plasmid, 2 restriction enzyme digests were run. BamHI (NEB, cat#R0136S) and PvuII (NEB, cat#R0151S), each with 2 cut sites on p30580, were used for confirmation (See Appendix B for digestion parameters). For the BamHI digestion, the plasmid was digested for 30 minutes at 37℃ and for the PvuII digestion, the plasmid was digested for 1 hour at 37℃. Each sample was then run on a 1% agarose gel at 130V for 30 minutes, comparing the results of both digestions to an undigested control for each sample. After successfully confirmation of purified plasmids, the confirmed plasmids were retransformed using the same protocol and harvested in bulk (10 ml) for further use.

Cassette Creation

For use in homologous recombination transformation, KanMx cassettes were generated via polymerase chain reaction (PCR) for each gene of interest. This was accomplished via primers with 20 bp of homology to the kanMx gene in p30580 and 60 bp of homology to regions directly upstream and downstream of the gene-of-interest, for the forward and reverse primers, respectively (Appendix B). After PCR amplification, products were confirmed on a 1% gel (130V, 30 minutes). The confirmed products were purified using a QIAquick PCR Purification Kit and quantified using a NanoDrop Spectrophotometer.

Super-High-Efficiency Yeast Transformation

BY4741 was inoculated in 5 ml liquid YPD media and incubated at 30℃ and 250 rpm overnight. After allowing colonies to grow overnight, the optical density (OD_{600}) was measured for each culture. 50 ml of fresh liquid YPD was then inoculated with enough culture to reach an OD_{600} of 0.1. The freshly inoculated YPD flasks were incubated at 30℃ and 250 rpm until

reaching an OD_{600} of around 0.5. In the meantime, 4 µg of purified kanMx cassette was assembled for each gene target in a final volume of 100 µl (in ddH_2O). Additionally, 20 µl of sheared salmon DNA (10 mg/ml) per sample was boiled for 8 minutes at 95℃, then quickly placed to chill for at least 5 minutes on ice.

After reaching the desired OD_{600}, samples were harvested via centrifugation at 2,000 rpm, RT, 10 minutes in a 50 ml tube. The supernatant was poured off and the cell was washed with 25 ml 1X LiTE. The washed cell was then centrifuged again at 4000 rpm, RT, 5 minutes, and the supernatant was removed via pipette. The residual cell pellet was resuspended in enough 1X LiTE to reach 30 optical density units (ODU) per ml (500 µl for OD_{600}=0.3, 1000 µl for OD_{600}=0.6). Then 100 µl of the newly concentrated cell solution was transferred to a new microfuge tube, to which 10 µl salmon sperm DNA and 100 µl assembled transformation DNA were added, vortexing the sample after each addition. This solution was incubated at RT for 30 minutes. After incubation, 790 µl PEG-LiOAc-DMSO solution was added to each tube and mixed gently, but thoroughly by pipette. This solution was incubated at RT for 30 minutes again. After incubation, the solution was heat shocked for 14 minutes at 42℃, mixing gently by pipette halfway through. The cells were then collected via centrifugation at 8,000 rpm, RT, 2 minutes and washed by gently resuspending in 250 µl $CaCl_2$ (5 mM). After 10 minutes, the cells were collected again via centrifugation at 8,000 rpm, RT, 2 minutes and resuspended in 1 ml YPD. These cell resuspensions were allowed to recover for 2 hours by incubation at 30℃, 200 rpm. After 2 hours, the cells were collected via centrifugation at 8,000 rpm, RT, 2 minutes. The supernatant was poured off and the cells were resuspended in residual media. The concentrated solutions were plated on YPD-Geneticin (G418, 200 µg/µl) plates (using glass beads (5-6 per plate) and incubated for 2-4 days at 30℃.

Heat Shock and Mock Heat Shock Harvest

Single colonies were inoculated in 5 ml YPD in 16 mm glass tubes. After around 8 hours of growth at 30℃ and 250 rpm, 200 ml YPD overnight cultures were inoculated, calculating the amount of starter culture necessary to reach an OD_{600}=0.8 at the desired time of harvest.

The following day, at least 30 minutes before the calculated harvest time, 100 ml flasks of YPD were placed at 55℃ to warm up prior to heat shock. Furthermore, 8 ml 37% formaldehyde and 92 ml dH_2O were placed on ice to cool. After reaching OD_{600}=0.7-0.9, 100 ml of the 200 ml culture was transferred to a clean, empty flask. One of the 100 ml flasks, labelled "Mock Heat Shock" (MHS) was crosslinked with 3 ml 37% formaldehyde and incubated at 25℃ for 15 minutes, after which point the crosslinking reaction was quenched with 6 ml 2.5 M glycine. Concurrently, the second 100 ml flask, labelled "Heat Shock" (HS), was heat shocked by adding 100 ml of 55℃ YPD, incubating at 37℃ for 6 minutes. After 6 minutes, the heat shock flask was crosslinked with 8 ml chilled 37% formaldehyde and 92 ml dH_2O, added simultaneously. The flask was moved to a 25℃ incubator for the crosslink reaction. After 15 minutes, the heat shock flask was quenched with 16 ml 2.5 M glycine. After 5 minutes, the quenched flasks were removed from the incubator and placed on ice until other samples were ready to be processed.

A cell pellet was collected via centrifugation for 7 minutes at 3,000 rpm and 4℃. This cell pellet, having removed the supernatant, was subsequently washed with 5 ml ST buffer + CPI. The washed cell pellet was collected again via centrifugation for 2 minutes at 10,000 rpm and 4℃. The cell pellet was washed once more with 1 ml ST + CPI, transferred to a 2 ml bead beating tube, and collected via centrifugation for 2 minutes at 10,000 rpm and 4℃. After removing the supernatant, the pellet was flash frozen in liquid nitrogen and stored at -80℃ until

further use.

Chromatin Preparation

Flash frozen cells were thawed at RT and moved to ice immediately after thawing. 1 ml of FA Lysis Buffer was added to each pellet, followed by 500 µl of 0.5 mm zirconia beads (BioSpec, cat#11079105z). Next, the tubes were placed in to a pre-chilled aluminum tube holder and beat for 3 minutes. After 3 minutes, the samples and aluminum block were placed at -20℃ for 5 minutes on a sheet of aluminum foil. The previous 2 steps were repeated 2 more times.

After bead beating, 13 mm glass tubes were set up for each sample. The bead beaten samples were transferred to the glass tubes by heating a needle to red hot with a Bunsen burner and poking a hole in the bottom of the sample tube before placing it into its designated glass tube. Then the needle was heated to red hot again and a hole was poked through the top of the sample tube. Liquid was collected, avoiding beads if possible, via centrifugation at ½ speed in a clinical centrifuge. The collected lysates were transferred into fresh 1.7 ml Eppendorf tubes.

The lysates were pelleted via centrifugation at for 3 minutes at 14,000 rpm and 4℃, pipetting off the supernatant. The pellets were then thoroughly resuspended in 600 µl FA Lysis Buffer before centrifuging again for 3 minutes at 14,000 rpm and 4℃. After aspirating off the supernatant, the chromatin pellets were resuspended in 800 µl FA Lysis Buffer. The resuspended pellets were transferred to 15 ml polystyrene tubes with 300 µl of 0.1 mm zirconia beads and sonicated in a Bioruptor Pico sonication device for 12 cycles, 15 seconds on and 30 seconds off. After sonication, the samples were transferred to fresh 1.7 ml Eppendorf tubes, avoiding beads as much as possible. 50 µl of each sample was transferred to a fresh tube for sonication check and the remainder of each sample was flash frozen and stored at -80℃ until further use.

PCIA Extraction and Sonication Check

The 50 µl set aside during chromatin preparation was reverse crosslinked at 65°C overnight with 175 µl 10 mM Tris-Cl pH 7.5, 225 µl 2X Proteinase K Buffer, and 2.5 µl Proteinase K (20 mg/ml). After reverse crosslinking overnight, DNA was extracted by adding 450 µl PCIA and vortexing thoroughly for 20 seconds. The samples were separated into two layers by centrifuging for 6 minutes at 14,000 rpm and RT. 400 µl of the top layer, containing extracted chromatin was carefully transferred via pipette to a new tube, avoiding contamination from the bottom layer. Then 42 µl of 3 M sodium acetate, 1 µl glycogen (20 mg/ml), and 1 ml 100% ethanol were mixed via vortex in each sample, which was incubated for 20 minutes at -80°C. After 20 minutes, a small white pellet was collected from each sample by centrifuging for 20 minutes at 14,000 rpm and 4°C. The supernatant was poured off and the pellet was washed carefully, without dislodging, with 500 µl 70% ethanol. Another small white pellet was collected via centrifugation for 5 minutes at 14,000 rpm and 4°C, this time carefully removing all liquid via pipette. The pellet was dried for about 20 minutes on the benchtop and resuspended in 30 µl TE + RNase, before allowing to incubate at 37°C for 2 hours. Next, all samples were run on a 1.8% agarose gel at 135 V for around 30 minutes to ensure that sonicated chromatin fragments were predominantly between 200-500 bp.

ChIP-exo 5.0

Day 1: Antibody Conjugation

15 µl of Protein G Mag Sepharose (GE Healthcare, cat#28-9513-79) beads were resuspended and transferred to a protein lo bind tube. After allowing beads to collect on a magnet, the bead buffer was removed by pipette. The beads were resuspended gently with 1 ml FA Lysis Buffer and set on the magnet again. After removing the FA Lysis Buffer, the beads

were resuspended in 1 ml FA Lysis Buffer again and set on ice. 5 µl of tRNA (10 mg/ml) was then added to the resuspended beads which were incubated spinning on a rototorque at 4℃ for 10 minutes. Concurrently, the desired antibodies were thawed on ice. After 10 minutes, the beads were placed on a magnet and FA Lysis Buffer and tRNA were removed from the beads, which were resuspended in 500 µl FA Lysis Buffer. Then 3 µg of antibody/sample was added to the resuspended beads, which were again placed to incubate spinning on a rototorque for 4-6 hours at 4℃.

After 4-6 hours, the conjugated beads were divided equally between all samples and placed on a magnet. Allowing the beads enough time to settle on the magnet, all liquid was removed and the beads were resuspended with 200 µl of chromatin per tube, resuspending thoroughly first. The strip tubes were placed to incubate spinning on a rototorque at 4℃ overnight.

Day 2: A-tailing, Adapter Ligation, Fill-in, Lambda Exonuclease, and Elution

After overnight incubation, the beads were removed from the rototorque and placed on ice until use. The samples were then washed with FA Lysis Buffer, NaCl Buffer 250, LiCl Buffer 250, and 10 mM Tris-HCl, pH 8 consecutively. Each wash was accomplished by placing the sample strip tube on a magnet, removing all liquid, and resuspending beads carefully with the designated wash buffer. It was important to not leave beads outside of buffer for extended periods of time to avoid drying.

The beads were kept resuspended in Tris-HCl until the A-tailing Reaction Mix (Appendix B) was prepared, at which point the samples were placed on a magnet and the Tris-HCl was removed by pipette. The A-tailing Reaction generated a 3’ A overhang and was set to shake for 30 minutes at 37℃ after carefully resuspending the beads. After incubation the beads were

washed with 10 mM Tris-HCl, pH 8 as previously done.

Next, 5’ adapter indexes were ligated and phosphorylated. This was done by adding the Adapter Ligation Reaction Mix to the beads after first removing Tris-HCl wash buffer. These indexes were important for sample differentiation during sequencing later. The adapter ligation reaction was incubated shaking for 1 hour at 25℃ and then washed consecutively with NaCl 250, LiCl 250, and Tris-HCl.

After adapter ligation, a fill-in reaction was completed with phi29 DNA polymerase to fill in the 3’ regions of the DNA fragments. This reaction was shaken for 20 minutes at 30℃, after which point the beads were washed in Tris-HCl again. Then, after removing all liquid from beads, the samples were resuspended in the Lambda Exonuclease Digest Mix and shaken for 30 minutes at 37℃. The lambda exonuclease, with 5’-3’ catalytic activity, was essential for achieving the high-resolution characteristic of ChIP-exo, digesting DNA to approximately 6 bp away from the crosslinked protein. Following the reaction, the beads were washed with Tris-HCl. After washing the beads and removing the liquid on a magnet, the beads were eluted and reverse crosslinked overnight in ChIP Elution Buffer and Proteinase K at 65℃.

Day 3: AMPure Purification, Adapter Ligation, and PCR

The following day, sample supernatants were eluted from their beads and transferred to a new strip tube to be purified using AMPure beads. Ensuring that the bead slurry was well mixed before addition to eluted samples, 72 µl of beads was added to each sample and mixed. After adding AMPure, the samples were incubated for 5 minutes on the benchtop and then placed on a magnet for 2 minutes. The beads were washed 2 times with chilled 70% ethanol before removing all liquid and letting the samples dry for 5 minutes. After drying, the samples were eluted in 21 µl ddH_2O and, after sitting for 2 minutes, transferred to new tubes.

A second ssDNA adapter ligation reaction was then carried out to attach another index to the DNA. This was carried out, shaking, at 25℃ for 1 hour, attaching one index (A-H) to each sample. Following, the samples were purified using AMPure beads as done before one more time, eluting the supernatant into a fresh strip tube at the end. The eluted DNA was PCR amplified for 24 cycles, removing 28 µl of sample for sequencing after 18 cycles. The remaining 12 µl of sample was used for gel confirmation of the library preparation.

Day 4: Gel Extraction

After first confirming a robust library signal between 200 and 500 bp on a 1% agarose gel using the samples PCR-amplified for 24 cycles, the 28 µl of sample removed after 18 cycles was gel purified on a 1% agarose gel (100 V, 1 hour) using a QIAquick Gel Extraction Kit. It was important to avoid overexposing the gel to UV to avoid mutating assembled libraries. This was accomplished by excising only the ladder and exposing that section to UV light, making a small cut at the 200 and 500 bp marks. The sample-containing parts of the gel were then aligned to these small cuts outside of the UV light, cut out, and carried through the QIAquick Gel Extraction Kit. All optional protocol steps for sequencing were performed and the final purified DNA was eluted in 30 µl of preheated (50℃) ddH_2O. After elution the samples were quantified using a Qubit Fluorometer (Invitrogen) and the included Qubit Assays protocol.

Data Analysis

Sequencing data was processed into heatmaps and composite plots using ScriptManager v0.12 (https://github.com/CEGRcode/scriptmanager/releases/tag/v0.12), a collection of Java scripts generated for the Pugh Lab data analysis pipeline. First, all raw sequencing files of type Bam, merged from two biological replicates, were mapped relative to a reference under the "Sequencing Read Analysis" tab and "Tag Pileup" script. Here, Bam files were sorted in order of

decreasing occupancy for the ribosomal protein, SAGA-dominated, and TFIID-dominated gene classes. Ribosomal protein genes encode the proteins that make up part of the complete composition of ribosomes. SAGA-dominated and TFIID-dominated gene classes each refer to a class of genes that are primarily mediated by a specific protein complex, either SAGA or TFIID. After sorting, these protein occupancy levels were centered around the transcription start site (TSS). Alternatively, a different set of analyses required that genes were sorted in order of increasing length and centered around the gene midpoint between the TSS and transcription end site (TES).

Files were mapped specifically with the following script parameters. The read to output was Read 1, selected to yield data with exonuclease resolution. Data was mapped without filtering and output strands were combined. A 6 bp tag shift was imposed on the data to account for the protein crosslink point. Tags were set to be equal, at 12 million tags, to ensure that all files had equivalent depth of sequencing. Bin size was maintained at 1 bp, a sliding window of 3 bp was set, and data was not smoothed. CDT files were important for heatmap generation, while OUT files were used to generate composite plots.

After mapping, outputted files of type CDT were loaded into the “Heat Map” script, under the “Figure Generation” tab. A contrast threshold was set at an absolute value of 3, while all other parameters were maintained.

Prism was used to generate composite plots from OUT files. The composite plots plotted number of sequencing reads at each gene locus within 1,000 bp to the TSS, comparing data between HS and MHS samples.

Chapter 3. Results

Binding Datasets are Consistent Across Biological Replicates

Final datasets were generated from two biological replicates (Figure 4A). Each biological replicate was processed identically, from cell harvest to ChIP-exo and sequencing. To ensure data was consistent and representative of each replicate, final datasets were merged datasets of two biological replicates (Figure 4B).

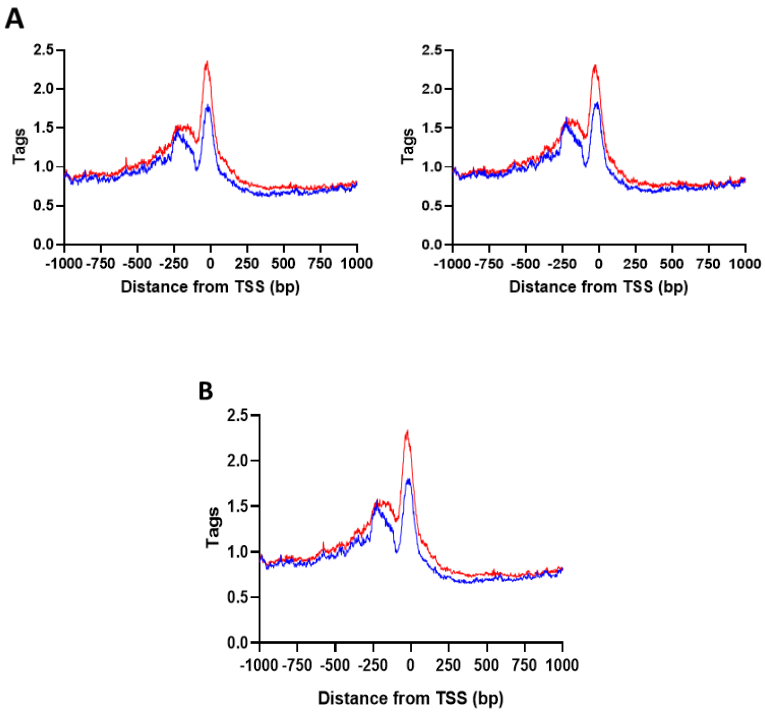


Figure 4. Composite Plots of Control TBP Biological Replicates

A) Biological replicates for BY4741 control data across the entire yeast genome, targeting TBP. Each graph measures binding occupancy of TBP (tags) centered around the TSS for HS (red) and MHS (blue) samples. B) Merged BY4741 control data closely resembles each biological replicate.

Stress-Induced PIC Assembly is Most Strongly Upregulated in SAGA-Mediated Genes

A heatmap is a useful tool to graphically represent matrix values using differential coloring or shading. Here, heatmaps were used to visually show protein occupancy levels across the yeast genome in greyscale, meaning highly occupied regions appear dark grey or black, while more sparsely occupied regions appear light grey or white. Each row in the heatmap is a gene in the yeast genome, separated into RP, SAGA-dominated, and TFIID-dominated genes. The occupancy data is then centered around a relevant location marker, like the gene-body midpoint or TSS.

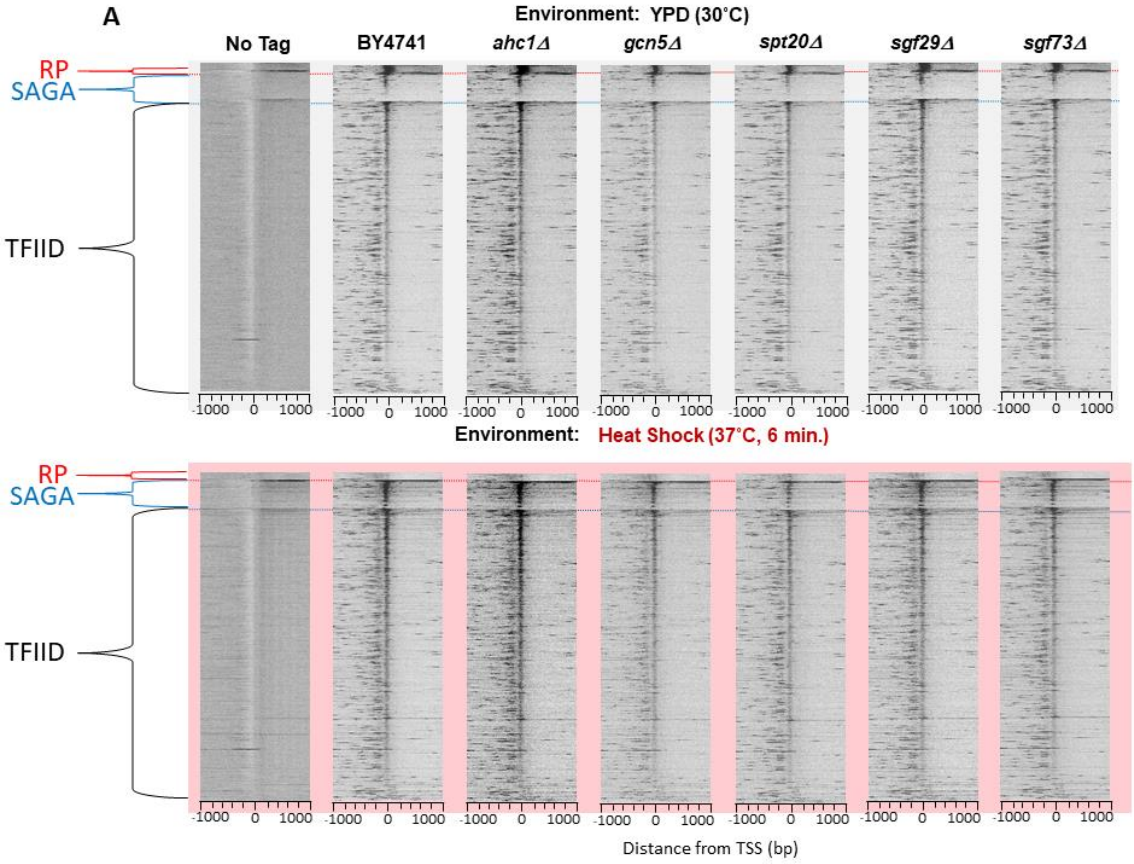
Figures 5A and 5B show a comprehensive summary of TBP and Pol II binding under all knockout conditions and each growth condition. RP genes, at the top of each heatmap and sorted from the top down by expression level, were consistently highly occupied under all conditions of normal growth, regardless of whether part of the SAGA complex was knocked out. Likewise, the most highly expressed TFIID-dominated genes were frequently occupied under normal growth conditions. However, SAGA-dominated genes were relatively inactive until stressed, at which point occupancy levels dramatically increased. Simultaneously, RP gene and TFIID-dominated gene occupancy decreased as the cells were stressed. These results confirm the role of SAGA-dominated genes in stress response, while asserting the RP and TFIID-dominated genes as more general housekeeping genes that are not as highly transcribed when the cell must respond to threatening environmental changes.

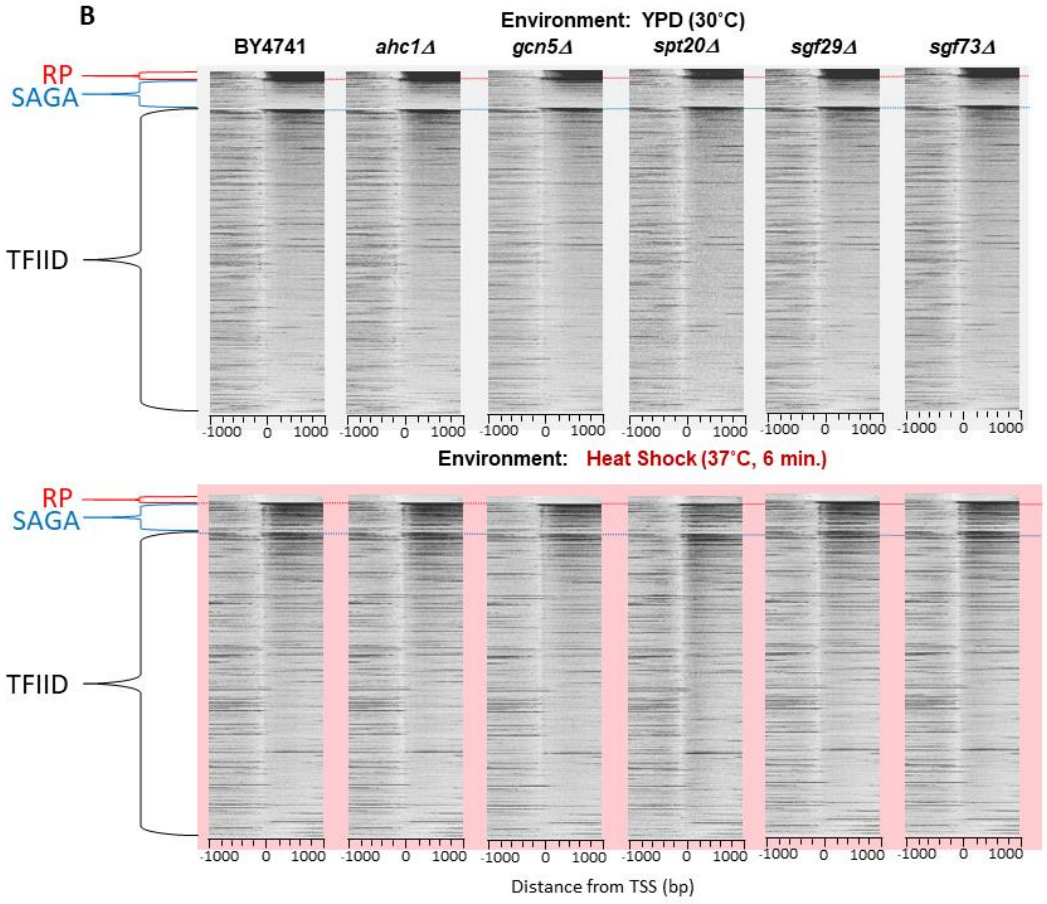
Notably TBP and Pol II show contrasting binding patterns in Figures 5A and 5B. Most heatmap occupancy for the TBP-targeted cells is located right next to the TSS with a sharply defined binding site. This pattern reflects TBP’s sequence-specific nature, recognizing and binding to the TATA box close to the TSS. Conversely, Pol II shows a dark, high occupancy

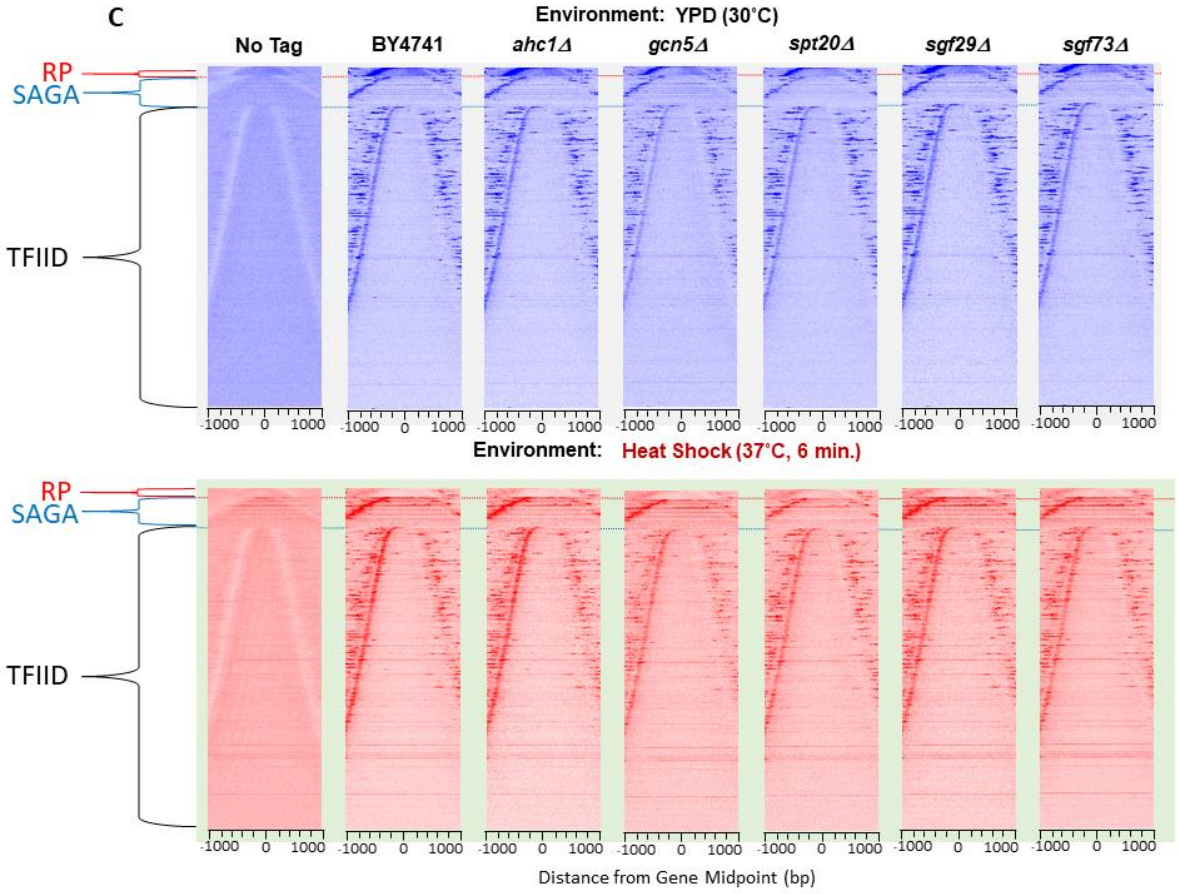
binding pattern even 1000 bp downstream from the TSS, reflecting Pol II's mobility along the gene during transcription. Rather than binding to one sequence and staying there during PIC assembly, like TBP, Pol II moves along the gene body and shows occupancy throughout.

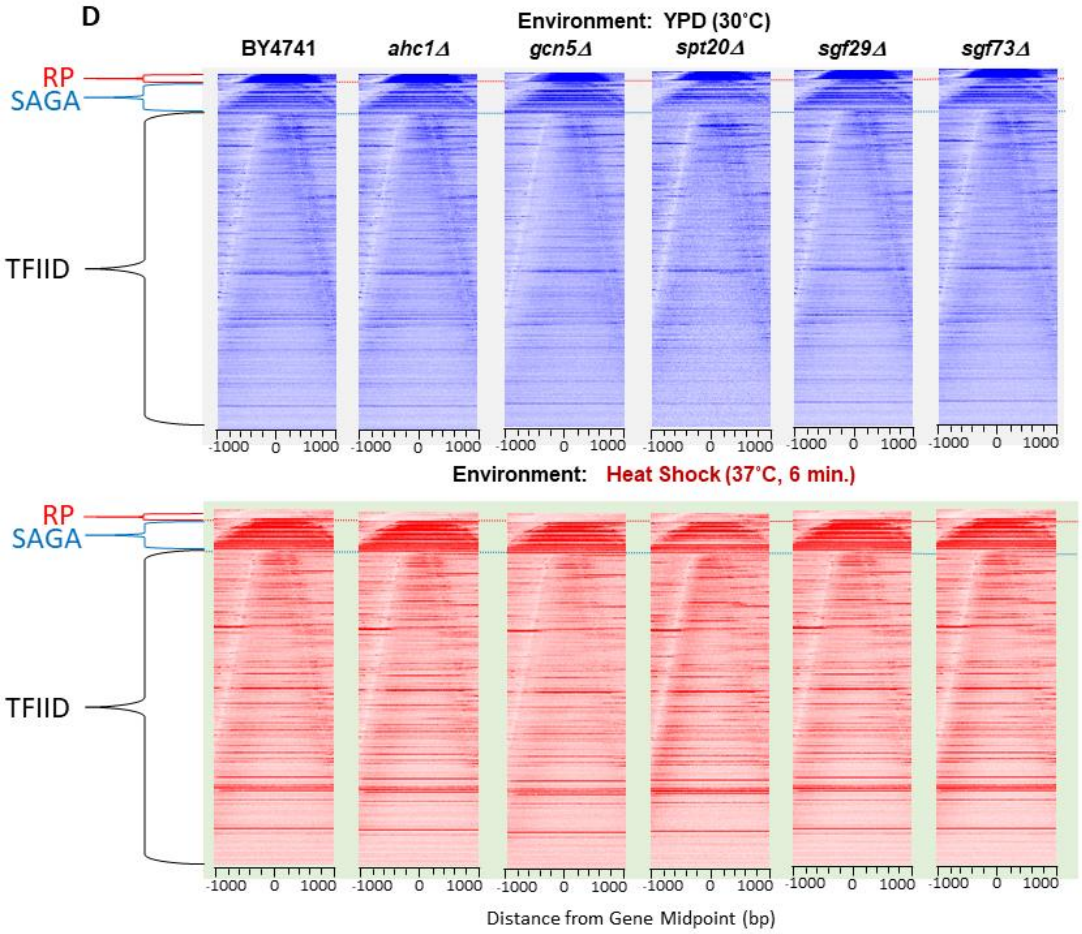
This distinct pattern is most noticeable in Figures 5C and D, which map TBP and Pol II binding, respectively, against gene midpoint. While the TBP heatmaps consistently show the highest signal directly to the left and right of the gene in a clear bell-shaped pattern, the Pol II heatmaps show an inverse effect, with the interior of the gene body shaded more darkly. This is evidence that TBP binds upstream or downstream of the gene, while Pol II binds the entire gene, travelling its length as it produces an mRNA transcript.

While SAGA is apparently most active in PIC recruitment under stress conditions, Figure 5E zooms in on its role across growth conditions in wild-type yeast. Binding occupancy in the most highly expressed SAGA-dominated genes near the top of the heatmap did not change drastically from normal growth to heat shock growth in the enhanced image, only mapped to SAGA-dominated genes. However, the less highly transcribed genes towards the bottom of the heatmap noticeably increased in occupancy, suggesting these genes were activated by the heat stress and are mostly dormant during optimal growth. The SAGA complex therefore has a role in transcription under both normal and stressed growth, but this role becomes more prominent during heat shock.


A
Environment: YPD (30°C)
No Tag
BY4741
ahc1Δ
gcn5Δ
spt20Δ
sgf29Δ
sgf73Δ
RP
SAGA
TFIID
-1000 0 1000
Environment: Heat Shock (37°C, 6 min.)
Distance from TSS (bp)


B
Environment: YPD (30°C)
BY4741
ahc1Δ
gcn5Δ
spt20Δ
sgf29Δ
sgf73Δ
RP
SAGA
TFIID
-1000 0 1000
Environment: Heat Shock (37°C, 6 min.)
RP
SAGA
TFIID
-1000 0 1000
Distance from TSS (bp)


C
Environment: YPD (30°C)
No Tag
BY4741
ahc1Δ
gcn5Δ
spt20Δ
sgf29Δ
sgf73Δ
RP
SAGA
TFIID
-1000 0 1000
Environment: Heat Shock (37°C, 6 min.)
Distance from Gene Midpoint (bp)


D
Environment: YPD (30˚C)
BY4741
ahc1Δ
gcn5Δ
spt20Δ
sgf29Δ
sgf73Δ
RP
SAGA
TFIID
-1000 0 1000
Environment: Heat Shock (37˚C, 6 min.)
Distance from Gene Midpoint (bp)

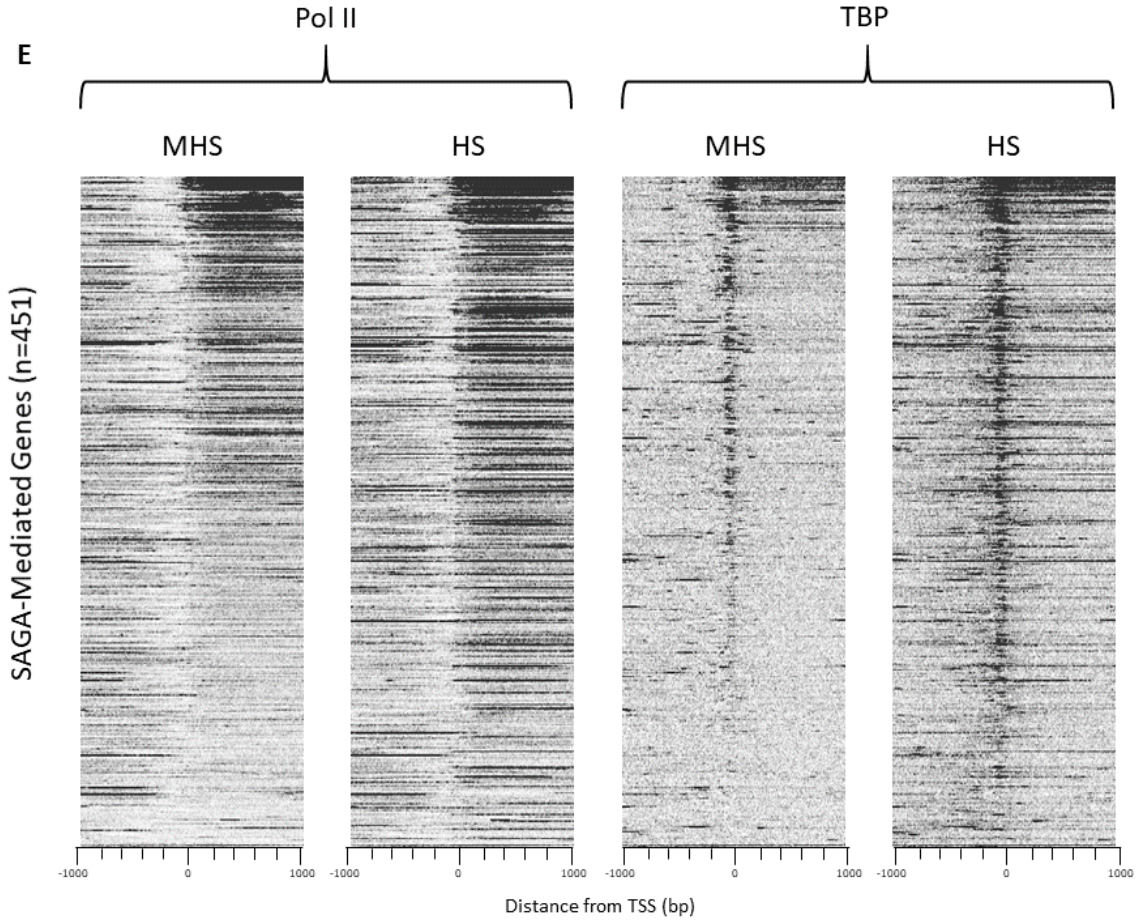


Figure 5. Genome Wide and SAGA-Mediated Recruitment of TBP and Pol II by the SAGA Complex Under Normal and Heat Shock Growth Conditions

A) TBP occupancy. Yeast cultures were grown under optimal conditions (top, 30°C) or under heat shock stress (bottom, 37°C for 6 minutes). The datasets are sorted in order of decreasing binding occupancy from top to bottom across three sets of genes: RP genes (n=130), SAGA-mediated genes (n=451), and TFIID-mediated genes (n=4,260) relative to the TSS. The No Tag heatmap, for MHS (top) and HS (bottom) samples, served as a negative control for both TBP and Pol II enrichment. With no epitope tag, these plots show only ChIP background noise. Negative control datasets were obtained from Pugh Lab master No Tag control datasets. Comparison of enrichment in MHS and HS datasets to the negative controls allowed for differentiation between signal and noise. B) Pol II occupancy. Yeast cultures were grown and datasets are sorted as described in A. C) TBP occupancy. Yeast cultures were grown as described in A. The datasets are sorted in order of increasing gene length, from top to bottom, across RP, SAGA-mediated, and TFIID-mediated genes. They are centered around the gene midpoint. D) Pol II occupancy. Yeast cultures were grown as described in A and datasets are sorted as described in D. E) Wild type yeast (BY4741) are shown mapped relative to only SAGA-mediated genes, sorted in order of decreasing binding occupancy from top to bottom.

Gcn5 and Spt20 Deletions Impair TBP Recruitment to the TSS

Composite plots, quantitatively comparing MHS and HS binding enrichment, are a valuable tool to directly assess the effects of an imposed cellular condition on protein binding. In Figure 6A, tag counts for TBP, in both wild-type and mutant yeast strains, were plotted centered around the TSS. In Gcn5 and Spt20-deficient SAGA complexes, TBP accumulated almost half as efficiently at and near the TSS as in the wild-type SAGA complex. This effect was present in both MHS and HS conditions but was more pronounced under stress. TBP's inefficient binding suggests that Gcn5 and Spt20 may play an especially critical role in SAGA's heat shock response, with the deletion strains recruiting much less effectively.

Interestingly, despite the marked decrease in TBP binding, Pol II binding was not as significantly impaired in MHS or HS in the deletion strains (Figure 6B). Pol II's ability to move throughout the gene open reading frame (ORF) appears to be slightly impaired far away from the TSS in the Spt20-deleted complex, but not as much as would be predicted from the drastic drop-off in TBP binding. TBP is traditionally required to recruit transcription machinery, including Pol II, so a decrease in TBP binding would typically decrease Pol II binding occupancies.

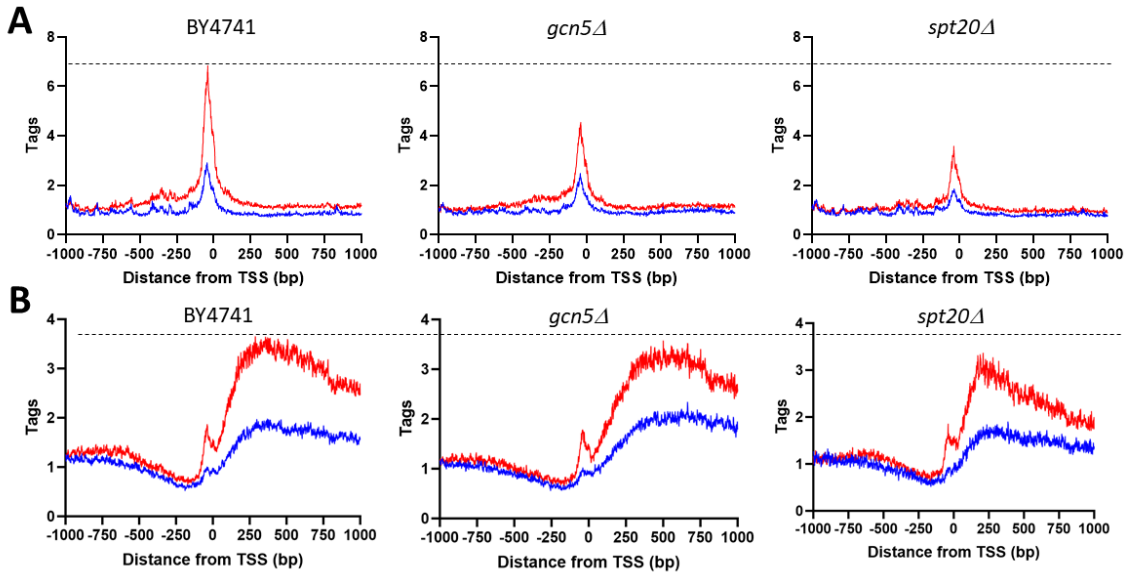


Figure 6. TBP and Pol II Recruitment by Gcn5 and Spt20-Deletion Strains

Composite plots comparing TBP (A) and Pol II (B) tag counts in MHS (blue) and HS (red) were generated for BY4741, *gcn5Δ*, and *spt20Δ* and grown as described in Figure 5. Each graph measures binding occupancy of TBP or Pol II (tags) in SAGA-dominated genes (n=451) centered around the TSS.

Ahc1 Deletions Do Not Affect PIC Assembly

SAGA and ADA share several subunits and may have overlapping roles. Therefore, since deletions in the SAGA complex resulted in reduced TBP and Pol II recruitment, I examined the effect of deleting Ahc1, a subunit unique to the ADA complex, on PIC assembly (Figure 7). Comparison of binding genome wide (Figure 7A, C) resulted in essentially no change in TBP or Pol II binding compared to control data (Figure 7E). This suggests several possibilities. Most simply, ADA may not have any role in PIC assembly. Alternatively, ADA may be dysfunctional, but SAGA may be compensating for deficiencies in ADA to maintain levels of TBP and Pol II binding. Another possibility is that ADA may have a role in PIC assembly, but the Ahc1

knockout may not have compromised the ADA complex's ability to recruit TBP and Pol II. However, in vitro studies show that Ahc1 maintains ADA complex structural integrity, so this third possibility is unlikely. Further studies examining potential ADA-mediated PIC assembly are necessary to fully characterize any analogous role in ADA to SAGA-mediated TBP and Pol II recruitment.

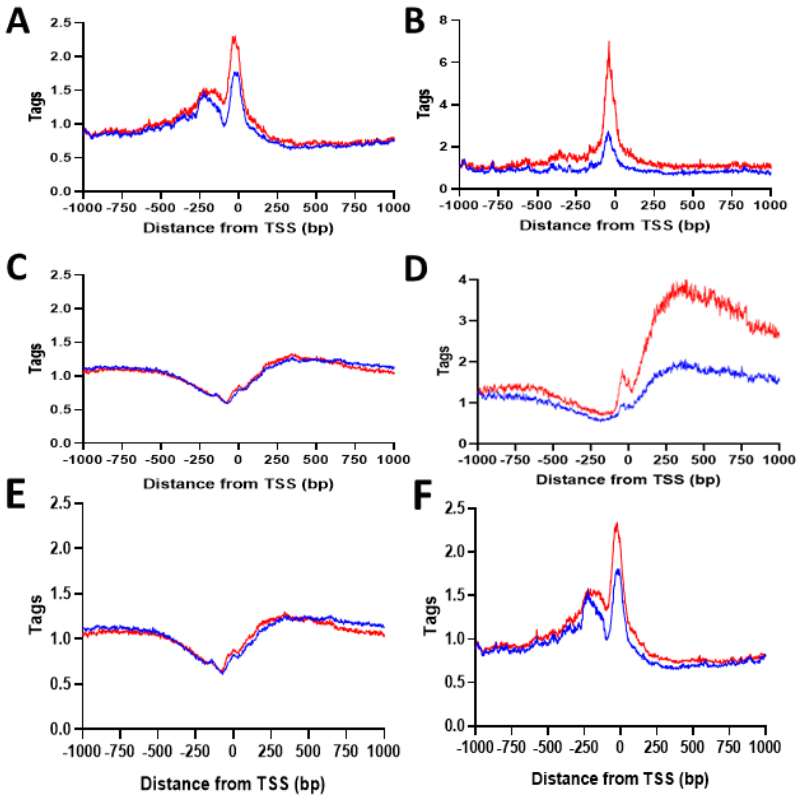


Figure 7. SAGA Response to an Ahc1-Deficient ADA Complex

Composite plots comparing the relative binding occupancies of TBP (A, B) and Pol II (C, D) for MHS (blue) and HS (red) samples. Plots A and C map occupancies relative to the TSS across the entire yeast genome. Plots B and D map only SAGA-dominated genes. E) BY4741 control data across all genes for Pol II. F) BY4741 control data across all genes for TBP.

Chapter 4. Discussion

The SAGA complex is a critical gene regulatory complex that is conserved across eukaryotic species. While well-studied since its discovery, SAGA remains incompletely characterized. This study aimed to address some information lapses in SAGA's role dealing with cellular stress, including its response to heat shock stress, subunit knockouts, and ADA complex deficiencies.

When exposed to heat shock stress, SAGA-dominated genes were consistently upregulated compared to normal growth conditions. This finding confirms results from previously cited research on SAGA and asserts SAGA as a stress-response complex. The other gene sets examined, RP genes and TFIID-mediated genes, showed significant drop offs in occupancy during heat shock stress, asserting that many non-SAGA-dominated genes are transcribed less often when under threatening cellular stress.

Subunit knockouts provided another proxy to examine how SAGA responds to stress conditions. Importantly, deficiencies in SAGA subunits like Gcn5 and Sg29 have been linked to disease development. Therefore, understanding the mechanisms behind SAGA's response to internal deletions remains an important clinical research topic. Here, Gcn5 and Spt20 were the most deleterious subunit deletions, showing an impaired ability to recruit TBP to the TSS compared to control complexes. Strikingly, however, Pol II recruitment was not impaired to the same degree, showing only a slight drop off in occupancy in genes far away from the TSS in *spt20Δ*. While initially surprising, there are several possible explanations for this observation. Most simply, Pol II may be more easily crosslinked than TBP, since it tends to bind to the gene longer. This could result in a higher signal even at the same biological binding frequency. Furthermore, transcription rate is slower in mutant cells, thus keeping Pol II on the gene body for

a longer time. This may have resulted in a higher tag count, despite diminished Pol II binding at the promoter due to the reduction in TBP recruitment. Another explanation that merits further investigation, but is outside the scope of this thesis, is that Pol II may be recruited by other factors than just TBP. Since Pol II typically depends on TBP binding to set PIC assembly into motion, eukaryotic species have been reported to be able to adapt to TBP deficiencies by substituting similar proteins to initiate PIC assembly, an adaptation that has been studied more in mammals than yeast[20]. Despite inability recruit TBP to site, TBP-associated factors (TAF(II)s) or TBP-like proteins may be able to compensate for an absence of TBP and maintain transcription levels. This insight may suggest an avenue for future research into which factors are recruited for PIC assembly instead of PIC, and what role SAGA may play in assembling them. Otherwise, comparison of wild-type and mutant RNA synthesis rates or crosslinking efficiency of TBP and Pol II may resolve the discrepancy.

ADA has several shared subunits with SAGA so ADA's potential role in PIC assembly and its relationship with SAGA was explored here. In Ahc1-deficient cells, binding profiles of TBP and Pol II showed no change compared to control cells, indicating that either SAGA may be helping compensate for a damaged ADA complex or that the ADA complex does not have a role in PIC assembly. Binding location of the SAGA complex itself was less explored here, since this study's scope aimed primarily to explore PIC assembly in relation to a stressed SAGA complex. However, further studies into where the SAGA complex itself bound may better define how SAGA and ADA may interact and communicate with each other. Furthermore, examining ADA's binding behavior or epigenetic activity with Ahc1 deletions may be useful to confirm whether the complex was significantly damaged. If Ahc1 knockouts did not significantly damage

the ADA complex, different subunit knockouts could be tested to determine SAGA's ability to compensate for ADA impairment. So far ADA's role in PIC assembly remains dubious.

Besides these two future directions, ongoing studies may aim to test how knockouts in the remaining SAGA subunits affect TBP and Pol II recruitment. Assessing each of SAGA's 20 subunits was outside the scope of this thesis, but complete characterization of internal stress in the SAGA complex may help pinpoint which subunits are most important for SAGA-related diseases, as well as SAGA's regulatory functions. Two other future studies may look at how complete depletion of the entire SAGA complex affects PIC assembly or how the cell's epigenomic profile changes with subunit deletions.

Apparently, Gcn5 and Spt20 have the most influence on TBP and Pol II binding out of the five subunits tested in this study. Likely, the other three subunits tested have little impact on TBP and Pol II recruitment. Gcn5, however, has HAT activity that is likely responsible for acetylating nucleosomes around the TATA box, promoting more efficient TBP binding to site. Additionally, Spt8 and Spt3, two other SAGA subunits, have been reported to interact directly with TBP in bringing it to the PIC[21]. It is possible that Spt20, in conjunction with Spt8 and Spt3, helps in bringing TBP to site. Therefore, absence of Spt20 would result in less efficient, but not a complete loss of, recruitment of TBP (Figure 8). Deletion of both Gcn5 and Spt20 would likely result in a further loss of TBP binding, with less efficient recruitment and binding of TBP at the TSS.

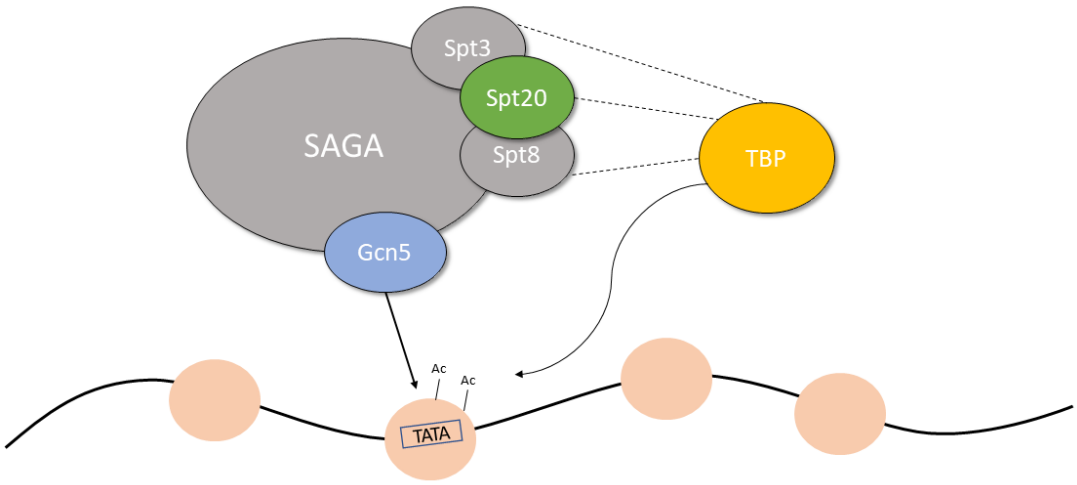


Figure 8. Gcn5 and Spt20-Mediated Recruitment of TBP

Gcn5 and Spt20-deficient yeast strains demonstrated the most significant reduction in TBP recruitment to the TSS. In this model of SAGA-mediated PIC assembly, Gcn5's HAT activity acetylates nucleosomes nearby the TATA box, resulting in more efficient binding of TBP. Loss of Gcn5 results in more tightly wound DNA and less efficient TBP binding. In collaboration with Spt3 and Spt8, Spt20 directly interacts with TBP, bringing it to site. Loss of Spt20 results in less efficient TBP binding, and a partial loss in TBP binding at the TSS[21].

While a preliminary and brief look into SAGA's stress-response mechanisms, this study begins to delve into SAGA's gene regulatory traits and puts forth future directions. Confirming SAGA's role in stress response *in vivo* suggests the applicability of previous *in vitro* studies to mammalian physiology, but more work must be done to fully characterize the SAGA complex in a mammalian system. Insights from these future studies will hopefully yield new information into the characteristics of a critical regulatory complex, and a significant player in disease development and prevention.

Appendix A

Reagents

LB-Ampicillin Plates (500 ml)			[Final]
Yeast Extract	2.5	g	
Bacto-Tryptone	5	g	
Sodium Chloride	5	g	
Agar	10	g	
Ampicillin (100 mg/ml)	500	µl	100 µg/ml
dH_2O			

Fill a 1 L beaker with 250 ml dH_2O and add yeast extract, bacto-tryptone, and sodium chloride one-at-a-time, spinning with a magnetic stir bar. After all solids are dissolved, bring the volume to 500 ml with dH_2O and add the agar. Stir briefly. Cap the bottle loosely and autoclave 20 minutes on liquid cycle. After allowing the solution to cool to around 55°C, add ampicillin, swirl bottle to mix, and pour plates. Allow the plates to cool completely to room temperature before storing at 4°C.

LB-Kanamycin Plates (500 ml) [Final]

Yeast Extract	2.5	g	
Bacto-Tryptone	5	g	
Sodium Chloride	5	g	
Agar	10	g	
Kanamycin (50 mg/ml)	500	µl	50 µg/ml
dH_2O			

Fill a 1 L beaker with 250 ml dH_2O and add yeast extract, bacto-tryptone, and sodium chloride one-at-a-time, spinning with a magnetic stir bar. After all solids are dissolved, bring the volume to 500 ml with dH_2O and add the agar. Stir briefly. Cap the bottle loosely and autoclave 20 minutes on liquid cycle. After allowing the solution to cool to around 55°C, add kanamycin, swirl bottle to mix, and pour plates. Allow the plates to cool completely to room temperature before storing at 4°C.

YPD (1 L) [Final]

Yeast Extract	10	g
Peptone	20	g
Dextrose	20	g
dH_2O		

Fill a 2 L beaker with 500 ml dH_2O and add reagents one-at-a-time, spinning with a magnetic stir bar. After all solids are dissolved, bring the volume to 1 L with dH_2O. Stir briefly. Cap the bottle loosely and autoclave 15 minutes on liquid cycle, placing bottle in a water bath to avoid burning the dextrose.

YPD Plates (1 L) [Final]

Yeast Extract	10	g
Peptone	20	g
Dextrose	20	g
Agar	20	g
dH_2O		

Fill a 2 L beaker with 500 ml dH_2O and add yeast extract, peptone, and dextrose one-at-a-time, spinning with a magnetic stir bar. After all solids are dissolved, bring the volume to 1 L with dH_2O and add the agar. Stir briefly. Cap the bottle loosely and autoclave 15 minutes on liquid cycle, placing bottle in a water bath to avoid burning the dextrose. After allowing the solution to cool to around 55°C, pour plates. Allow the plates to cool completely to room temperature before storing at 4°C.

Tris pH 7.5 (1 L) [Final]

Trizma	121.2	g	1 M
1 N HCl	800	ml	
ddH_2O			

Bring final volume to 1 L with ddH_2O and adjust pH with HCl as needed. Filter sterilize and store at RT.

0.5 M EDTA, pH 8.0 (500 ml)			[Final]
EDTA	93.1	g	0.5 M
NaOH Pellets	~10	g	
ddH_2O			

Fill a 1 L beaker with 350 ml dH_2O and add the EDTA, spinning with a magnetic stir bar. Add NaOH pellets one-at-a-time until the pH reaches 8.0, then bring the final volume to 500 ml with ddH_2O. Filter sterilize and store at RT.

10X LiTE (250 ml)			[Final]
Lithium Acetate	25.5	g	1 M
Tris pH 7.5	25	ml	0.1 M
EDTA	5	ml	10 mM
ddH_2O			

Bring the final volume to 250 ml with ddH_2O and filter sterilize. Store at RT. Dilute to 1X LiTE with ddH_2O on the day of transformation.

50% PEG 3350 (50 ml)			[Final]
PEG 3350	25	g	50%
ddH_2O			

Stir the solution with a magnetic stirrer, ensuring that all PEG 3350 dissolves since it is not very soluble in water. Bring the final volume to 50 ml with ddH_2O and filter sterilize. Store at RT in a dark place.

PEG-LiOAc-DMSO (790 µl per sample)			[Final]
50% PEG 3350	600	µl	30%
10X LiTE	90	µl	1X
DMSO	100	µl	10%

Prepare on the day of the transformation, 790 µl per sample. Mix by inverting.

2.5 M $CaCl_2$ (250 ml)			[Final]
$CaCl_2$	110.98	g	2.5 M
ddH_2O			

Bring final volume to 250 ml with ddH_2O, filter sterilize, and store at 4°C. On the day of transformation, dilute to 5 mM.

cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail (CPI) (1 ml)		
CPI Pellet	1	
ddH_2O	1	ml

Completely dissolve 1 CPI pellet in ddH_2O by vortexing.

ST Buffer (1 L)			[Final]
1 M Tris-HCl pH 7.5	10	ml	10 mM
5 M NaCl	20	ml	100 mM
ddH_2O			

Bring the final volume to 1 L with ddH_2O, filter sterilize, and store at 4°C. On the day of harvest, add CPI (2 µl/ml) to the solution and invert to mix.

2.5 M Glycine (500 ml)			[Final]
Glycine	93.75	g	2.5 M
ddH_2O			

Heat the ddH_2O on a hot plate before adding the glycine. Bring the final volume to 500 ml with ddH_2O, filter sterilize, and store at RT.

1 M Hepes/KOH, pH 7.5 (1 L)			[Final]
1 M HEPES	238.4	g	1 M
1 N KOH	500	ml	
ddH_2O			

Adjust pH with KOH as needed. Bring the final volume to 1 L with ddH_2O, filter sterilize, and store at 4°C.

FA Lysis Buffer (500 ml)			[Final]
1 M Hepes/KOH, pH 7.5	25	ml	50 mM
5M NaCl	15	ml	150 mM
0.5 M EDTA	2	ml	2 mM
10% NaDeoxycholate	5	ml	0.1%
25% Triton X-100	20	ml	1%
ddH_2O			

Bring the final volume to 500 ml with ddH_2O, filter sterilize, and store at 4°C. On the day of use, at CPI (2 µl/ml) and mix by inverting.

2X Proteinase K Buffer (50 ml)			[Final]
1 M Tris-HCl, pH 7.5	2	ml	40 mM
0.5 M EDTA	4	ml	40 mM
20% SDS	5	ml	2%
ddH_2O			

Bring the final volume to 50 ml with ddH_2O, filter sterilize, and store at RT.

Appendix B

Reaction Specifications

Restriction Digests

BamHI			[Final]
p30580 Plasmid	500	µg	10 µg/µl
BamHI	0.5	ml	
10X Cutsmart	5	µl	1X
dH_2O			

Bring the reaction mixture to 50 µl with dH_2O and incubate at 37℃ for 30 minutes.

PvuII			[Final]
p30580 Plasmid	500	µg	10 µg/µl
PvuII	0.5	ml	
10X Cutsmart	5	µl	1X
dH_2O			

Bring the reaction mixture to 50 µl with dH_2O and incubate at 37℃ for 30 1 hour.

Forward Primers for Homologous Recombination Gene Knockouts

Gene Target	Sequence
Ahc1	ttcactactcttcgggctcacgcttctcatccaacactttgtgtatatgtccatctcctcATGGGTAAGGAAAAGACTC
Gcn5	agggaagaccgtgagccgcccaaaagtcttcagttaactcaggttcgtattctacattagATGGGTAAGGAAAAGACTC
Spt20	agccttcaagaggattaggaaggaatagttacggttaatttgcgcctatatatttcagggATGGGTAAGGAAAAGACTC
Sgf29	ggtggcttctctaagtgggggggagtttttcacagcaaaacacacggtcacctttcttattATGGGTAAGGAAAAGACTC
Sgf73	aaaccaaaaaaataagaatagtgaacacacaagagaagcgcaaaagagtaaagagctaaaATGGGTAAGGAAAAGACTC

Lowercase letters are homologous sequences to regions upstream of the gene of interest.

Uppercase letters are homologous sequences to the kanMx gene.

Reverse Primers for Homologous Recombination Gene Knockouts

Gene Target	Sequence
Ahc1	attaggtcagaaaaccaacagaatattatattacgtaatttacttatttatatgtgtgtaTTAGAAAAACTCATCGAGC
Gcn5	cgtactaaacatttatttcttcttcgaaaggaatagtagcggaaaagcttcttctacgcaTTAGAAAAACTCATCGAGC
Spt20	tatatatatatatatatatatatatatatataaggaatgataactctatttaagtagaTTAGAAAAACTCATCGAGC
Sgf29	aataatgacaaaatgccatgagaagatcttatgatatgtagtaaatgttaaccaccattgTTAGAAAAACTCATCGAGC
Sgf73	tcggcgaataatttttttattactcacttcgtgaacatgctggataacgtgcatgattcaaTTAGAAAAACTCATCGAGC

Lowercase letters are homologous sequences to regions downstream of the gene of interest.

Uppercase letters are homologous sequences to the kanMx gene.

PCR Parameters for KanMx Cassette with Overhang

			[Final]
P30580 Plasmid	0.5	µl	30.53 ng
5x Phusion HF Buffer	10	µl	1X
dNTPs (3mM)	3.5	µl	0.245 mM
Forward Primer (10 µM)	1	µl	0.2 µM
Reverse Primer (10 µM)	1	µl	0.2 µM
Phusion HF/HS	1	µl	
ddH_2O	33	µl	

Phusion HF/HS from Thermo Fisher Scientific (cat#F-549L).

PCR Cycles (30)

1. 98℃ 1 min
2. 98℃ 15 sec |
3. 55℃ 20 sec | x10
4. 72℃ 1 min |

5. 98℃ 15 sec | x20
6. 72℃ 80 sec |

7. 72℃ 5 min
8. 4℃

ChIP-exo Reactions

Wash Buffers (150 µl/sample)

Reagent	Minimal Volume Needed per Sample (ml)	Total Volume (ml)	CPI (2 ul CPI/ml Buffer)
FA Lysis Buffer	0.15	10	20
NaCl Buffer 250	0.3	10	20
LiCl Buffer 250	0.3	10	20
Reagent	**Minimal Volume Needed per Sample (ml)**	**Total Volume (ml)**	**25% Triton-X (0.4 ul/ml Buffer)**
10 mM Tris-HCl, pH 8	1	30	12

A-tailing Reaction (40 µl/sample)

Reagent	Volume per Sample (µl)	[Final]	Total Volume (µl)
Water	30.3	-	799.9
10X NEBuffer 2	4.0	1X	105.6
3 mM dATP	2.7	200 µM	70.5
Klenow Frag, -exo (5 U/µl)	3.0	15 U	79.2
Total	40.0		1055.2

Adapter Ligation (40 µl/sample)

Reagent	Volume per Sample (µl)	[Final]	Total Volume (µl)
Water	7.0	-	184.8
2X Rapid Ligation Buffer	20.0	1X	528.0
T4 Ligase (600 U/µl)	2.0	1200 U	52.8
T4 PNK (10 U/µl)	1.0	10 U	26.4
Total	30.0		765.6
ExB2_i##/ExA2B (1.5µM)	10.0	15 pmol	

Dilute the index of choice, ExB2_i##/ExA2B (1.5µM), 1:10 in ddH_2O prior to use. 8 samples can have the same first index (1-18). Each of the 8 samples should have a unique second index (A-H), ExB1-iX/ExB1B-5N (3µM), diluted 1:5 prior to use.

Fill-in Reaction (40 µl/sample)

Reagent	Volume per Pool (µl)	[Final]	Total Volume (µl)
Water	24.6	-	649.4
10X BSA (1 mg/ml)	8.0	200 µg/ml	211.2
10X phi29 rxn buffer	4.0	1X	105.6
3mM dNTPs	2.4	180 µM	63.4
phi29 pol (10 U/ul)	1.0	10 U	26.4
Total	40.0		1056

Lambda Exonuclease Digest (40 µl/sample)

Reagent	Volume per Sample (µl)	[Final]	Total Volume (µl)
Water	31.6	-	834.2
10X Lamda Exo Rxn Buffer	4.0	1X	105.6
10% Triton-X	0.4	0.10%	10.6
DMSO	2.0	5%	52.8
Lambda Exonuclease (5 U/µl)	2.0	10 U	52.8
Total	40.0		1056

Elution and Reverse Crosslink (40 µl/sample)

Reagent	Minimal Volume Needed per Sample (µl)	Total Volume (µl)
ChIP Elution Buffer	40	1056
Proteinase K (20 mg/ml)	1.5	39.6

AMPure Purification (21 µl/sample)

Reagent	Minimal Volume Needed per Sample (µl)	Total Volume (ml)
AMPure Beads	72	2
70% Ethanol	400	10

PCR

Reagent	Volume per Sample (µl)	[Final]	Total Volume (µl)
Water	6.8	-	193.6
5X Phusion HF Buffer	8.0	1X	228.8
3mM dNTPs	2.67	200 µM each	76.4
P1.3 primer (20 µM)	0.80	500 nM	22.9
P2.1 primer (20 µM)	0.80	500 nM	22.9
Phusion DNA pol (2 U/µl)	1.0	2 U	28.6
Total	20.0		573.1
Resuspended DNA	20.0		

Remove 28 µl of sample after 18 cycles and allow the rest to run to a total of 24 cycles for use in gel confirmation of library prep.

PCR Cycles (24)

9. 94°C 5 min
10. 94°C 20 sec |
11. 50°C 1 min | x18
12. 72°C 1 min |

13. 4°C PAUSE
14. 94°C 2 min

15. 94°C 20 sec |
16. 50°C 1 min | x7
17. 72°C 1 min |

18. 72°C 5 min
19. 4°C

BIBLIOGRAPHY

[1] Spedale, G., Timmers, H. T., and Pijnappel, W. W. (2012) ATAC-king the complexity of SAGA during evolution, *Genes Dev 26*, 527-541.

[2] Eberharter, A., Sterner, D. E., Schieltz, D., Hassan, A., Yates, J. R., Berger, S. L., and Workman, J. L. (1999) The ADA complex is a distinct histone acetyltransferase complex in Saccharomyces cerevisiae, *Mol Cell Biol 19*, 6621-6631.

[3] Han, Y., Luo, J., Ranish, J., and Hahn, S. (2014) Architecture of the Saccharomyces cerevisiae SAGA transcription coactivator complex, *Embo j 33*, 2534-2546.

[4] Huisinga, K. L., and Pugh, B. F. (2004) A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae, *Mol Cell 13*, 573-585.

[5] Liu, W., Li, L., Ye, H., Chen, H., Shen, W., Zhong, Y., Tian, T., and He, H. (2017) From Saccharomyces cerevisiae to human: The important gene co-expression modules, *Biomed Rep 7*, 153-158.

[6] Helmlinger, D., and Tora, L. (2017) Sharing the SAGA, *Trends Biochem Sci 42*, 850-861.

[7] Lee, K. K., Sardiu, M. E., Swanson, S. K., Gilmore, J. M., Torok, M., Grant, P. A., Florens, L., Workman, J. L., and Washburn, M. P. (2011) Combinatorial depletion analysis to assemble the network architecture of the SAGA and ADA chromatin remodeling complexes, *Mol Syst Biol 7*, 503.

[8] Bonnet, J., Wang, C. Y., Baptista, T., Vincent, S. D., Hsiao, W. C., Stierle, M., Kao, C. F., Tora, L., and Devys, D. (2014) The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription, *Genes Dev 28*, 1999-2012.

[9] Durant, M., and Pugh, B. F. (2006) Genome-wide relationships between TAF1 and histone acetyltransferases in Saccharomyces cerevisiae, *Mol Cell Biol 26*, 2791-2802.

[10] Wang, L., and Dent, S. Y. (2014) Functions of SAGA in development and disease, *Epigenomics 6*, 329-339.

[11] Koutelou, E., Hirsch, C. L., and Dent, S. Y. (2010) Multiple faces of the SAGA complex, *Curr Opin Cell Biol 22*, 374-382.

[12] Kremer, S. B., and Gross, D. S. (2009) SAGA and Rpd3 chromatin modification complexes dynamically regulate heat shock gene structure and expression, *J Biol Chem 284*, 32914-32931.

[13] Vinayachandran, V., Reja, R., Rossi, M. J., Park, B., Rieber, L., Mittal, C., Mahony, S., and Pugh, B. F. (2018) Widespread and precise reprogramming of yeast protein-genome interactions in response to heat shock, *Genome Res*.

[14] Kassem, S., Villanyi, Z., and Collart, M. A. (2017) Not5-dependent co-translational assembly of Ada2 and Spt20 is essential for functional integrity of SAGA, *Nucleic Acids Res 45*, 7539.

[15] Bian, C., Xu, C., Ruan, J., Lee, K. K., Burke, T. L., Tempel, W., Barsyte, D., Li, J., Wu, M., Zhou, B. O., Fleharty, B. E., Paulson, A., Allali-Hassani, A., Zhou, J. Q., Mer, G., Grant, P. A., Workman, J. L., Zang, J., and Min, J. (2011) Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation, *Embo j 30*, 2829-2842.

[16] Yan, M., and Wolberger, C. (2015) Uncovering the role of Sgf73 in maintaining SAGA deubiquitinating module structure and activity, *J Mol Biol 427*, 1765-1778.

[17] Solomon, M. J., and Varshavsky, A. (1985) Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures, *Proc Natl Acad Sci U S A 82*, 6470-6474.

[18] Rhee, H. S., and Pugh, B. F. (2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution, *Cell 147*, 1408-1419.

[19] Rossi, M. J., Lai, W. K. M., and Pugh, B. F. (2018) Simplified ChIP-exo assays, *Nat Commun 9*, 2842.

[20] Wieczorek, E., Brand, M., Jacq, X., and Tora, L. (1998) Function of TAF(II)-containing complex without TBP in transcription by RNA polymerase II, *Nature 393*, 187-191.

[21] Sterner, D. E., Grant, P. A., Roberts, S. M., Duggan, L. J., Belotserkovskaya, R., Pacella, L. A., Winston, F., Workman, J. L., and Berger, S. L. (1999) Functional organization of the yeast SAGA complex: distinct components involved in structural integrity, nucleosome acetylation, and TATA-binding protein interaction, *Mol Cell Biol 19*, 86-98.

Matthew Driban

mdriban75@gmail.com

EDUCATION

The Pennsylvania State University (PSU), University Park, PA

Eberly College of Science | Bachelor of Science in Biochemistry and Molecular Biology

Schreyer Honors College | Phi Beta Kappa

RESEARCH EXPERIENCE

Dr. Frank Pugh's Biochemistry and Molecular Biology Lab (PSU), University Park, PA

Undergraduate Researcher (Jun 2015 – Present)

- Conducts research in protein-DNA interaction and transcription factor binding in E. coli and yeast
- Trains and mentors new undergraduate researchers
- Uses bacterial culture harvest, cell lysis and sonication, electrophoresis, mutation, and ChIP-exo protocols
- Analyzes results using Unix, GeneTrack, and scripts in Python and Java

Dr. Christiane Zweier's Genetics Lab (Friedrich-Alexander Universität Erlangen-Nürnberg), Erlangen, Germany

Research Intern, DAAD-RISE (May 2018 – Aug 2018)

- Studied the genetic bases for Borjeson-Forssman-Lehmann syndrome, a neurodevelopmental disorder
- Used mammalian cell culture, immunofluorescence, Western Blot, and luciferase protocols

Dr. Peter Aplan's Leukemia Biology Lab (National Institutes of Health), Bethesda, MD

Research Intern, National Institutes of Health Summer Internship Program (Jun 2017 – Aug 2017)

- Examined T cell receptor beta clonality in genetically engineered leukemia
- Used mammalian organ harvest, DNA/RNA isolation, PCR, gel electrophoresis, and Sanger sequencing protocols
- Analyzed results using Sequencher and BLAST

Dr. Magnus Ingelman-Sundberg's Pharmacogenetics Lab (KI), Stockholm, Sweden

Research Intern, Euroscholars Program (Jan 2017 – Jun 2017)

- Researched 3D hepatocyte spheroids to better understand drug-dependent enzyme induction and adverse drug reactions in humans
- Used mammalian cell culture, RNA isolation, cDNA preparation, and Real-Time PCR protocols

RESEARCH PRESENTATIONS

- Driban M, Fliedner A, Zweier C. (2018) Mechanistic Determination of PHf6's Role in Borjeson-Forssman-Lehmann Syndrome. DAAD-RISE Conference, Heidelberg University, Heidelberg, Germany. (Oral Presentation)
- Driban M, Yin M, Aplan P. (2017) T-cell Receptor Beta Rearrangements Determine the Clonality of Genetically Engineered Leukemia. National Institutes of Health Summer Poster Day, Bethesda, Maryland. (Poster Presentation)

- Driban M, Hendriks D, Ingelman-Sundberg M. (2017) Evaluation of 3D Hepatocyte Spheroids for the Mechanistic Understanding of Drug Dependent Enzyme Induction. Euroscholars Midstay Conference, KU Leuven, Leuven, Belgium. (Oral Presentation, 1st place for best research presentation in all categories)

AWARDS AND HONORS

- Dean's List, 2015-Present
- Schreyer Academic Excellence Scholarship, 2015-Present
- DAAD-RISE, 2018
- Louis A. Martarano Grant, 2017
- Schreyer International Study Award, 2017
- Euroscholar, Karolinska Institutet, 2017
- Euroscholars Midstay Conference, 1st Place Oral Presentation, 2017

VOLUNTEER EXPERIENCE

Mount Nittany Medical Center (University Park, PA)

Patient Floor, Education Department (May 2014 – Present)

- Assists non-surgical doctors of multiple specialties with day-to-day tasks on patient floors
- Worked in the hospital education department to aid in employee training updates

Remote Area Medical, University Park, PA

Doctor Outreach Chair (Feb 2018 – Present)

- Helps with the first leadership team of a new Penn State student organization, Remote Area Medical, designed to provide healthcare to low-income and underserved rural communities
- Worked at patient registration at a Remote Area Medical weekend clinic in Ashtabula, OH and as a clinic guide in Charleston, WV

TEACHING EXPERIENCE

Eberly College of Science, University Park, PA

Learning Assistant (Jan 2018 – Present)

- Taught a supplementary problem-solving class for students enrolled in introductory molecular biology courses (BMB 398)
- Took a science pedagogy course and had weekly interactions with the course professor to supplement in-class teaching

State College Area School District, State College, PA

SCASD Tutoring Service (Sep 2017 – Present)

- Tutors middle school and high school students in Math for 1 hour each week

LEADERSHIP EXPERIENCE

Presidential Leadership Academy, University Park, PA

Theta Class (Mar 2016 – Present)

- Participates as 1 of 30 students from a pool of over 300 applicants in a 3-year academic and extra-curricular leadership development program instructed by the University President and the Dean of the Schreyer Honors College
- Drafted and presented a 60-page policy, adopted by Penn State, to better utilize technology to improve college student mental health

Science Lion Pride, University Park, PA

Member (Jan 2016 – Present)

- Engages in science outreach through prospective high school student recruitment, alumni engagement, and service
- Recruits new students to the Eberly College of Science by leading personalized individual and group tours through Penn State science buildings
- Facilitates Eberly events including poster exhibitions and alumni meals