Title

Exploring the Regulatory Role of Major Yeast Histone Acetyltransferase Gcn5 in pre-mRNA Splicing Genome-wide

Author

Okonkwo, Shawntel

Publication Date

2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Exploring the Regulatory Role of Major Yeast Histone Acetyltransferase Gcn5 in pre-mRNA Splicing Genome-wide

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy in Molecular Biology

by

Shawntel Udoka Okonkwo

2020

ABSTRACT OF THE DISSERTATION

Exploring the Regulatory Role of Major Yeast Histone Acetyltransferase Gcn5 in pre-mRNA Splicing Genome-wide

by

Shawntel Udoka Okonkwo

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2020

Professor Gregory Payne, Chair

The step-wise assembly of the spliceosome onto pre-mRNA occurs co-transcriptionally, while the nascent transcript is synthesized from RNA Polymerase II. The mechanisms underlying the molecular coupling and coordination of these co-transcriptional reactions have not been thoroughly elucidated in the context of the dynamic chromatin environment. The Johnson Lab previously discovered that the major yeast histone acetyltransferase (HAT) Gcn5 demonstrates genetic interactions with *MSL1* and *LEA1*—which encode two core U2 small nuclear ribonucleoprotein particle (snRNP) components of the spliceosome. Additionally, the lab observed that Gcn5 HAT activity is required for proper co-transcriptional recruitment of the spliceosome in yeast where U2 snRNP association with the pre-mRNA and subsequent spliceosomal rearrangements are sensitive to Gcn5-dependent acetylation. While Gcn5-

dependent histone acetylation is important for the fate of spliceosomal rearrangements of these two proteins, the genome-wide implications and overall mechanism underlying this result is not yet clear. Here, I employed RNA-seq in *S. cerevisiae* to identify mechanistic insights into how Gcn5-dependent histone acetylation can affect pre-mRNA splicing. I prepared libraries from the following yeast strains: wildtype, *gcn5Δ* and *H3Δ9-16* (deletion of residues 9-16 from histone H3 N-terminal tail) such that genes for which splicing is similarly affected in both *gcn5Δ* and *H3Δ9-16* will represent a defect specific to the absence of Gcn5-histone acetyltransferase (HAT) activity. Surprisingly, the results support a role for the Gcn5-HAT in decreasing gene expression of all intron-containing ribosomal protein genes (IC-RPGs). Consequently, *gcn5Δ* and *H3Δ9-16* improved the splicing efficiency of a subset of IC-non-RPGs in a manner regulated by the competition of IC-RPGs for limited spliceosomes. Lastly, *gcn5Δ* and *H3Δ9-16* resulted in the decreased splicing efficiency for another subset of IC-non-RPGs. With publicly available ChIP-seq and MNase-seq data, I show that despite the increased availability of limited spliceosomes via down-regulation of IC-RPGs, *gcn5Δ* and *H3Δ9-16* dependent splicing outcomes of IC-non-RPGs are distinguished by differences in RNAPII, H3K9ac and MNase enrichment profiles. In this thesis, I uncover multiple effects of the *gcn5Δ* and *H3Δ9-16* on pre-mRNA splicing genome-wide--encompassing mechanisms regarding the economics of limited spliceosome availability and distinct chromatin landscapes directly regulating splicing outcomes.

The dissertation of Shawntel Udoka Okonkwo is approved.

Michael Carey

Albert Courey

Jeffrey Long

Gregory Payne, Committee Chair

University of California, Los Angeles

2020

DEDICATION

I dedicate this dissertation to my maternal grandmother, Rosaline Amodo Nwanekolu Anaduaka, my younger brother, Chinedu Okonkwo and my older brother Desmond-Tutu Okonkwo. I also dedicate this dissertation to my late friend, fellow UCLA PhD student and brilliant colleague, Ms. Taylor Brown. May her gentle soul rest in perfect peace.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

"Ndi ora n'azu nwa [It takes a village to raise a child]" - Igbo proverb

"'For I know the plans I have for you,' declares the Lord. 'Plans to prosper to you and not harm you, plans to give you a future and a hope'" - Jeremiah 29:11

I would like to thank my funding sources: the National Science Foundation (NSF) Graduate Research Fellowship Program, the UCLA Graduate Division's Competitive Edge Fellowship Program, NSF-LSAMP Bridge to the Doctorate Fellowship Program, Eugene V. Cota-Robles Fellowship and the UCLA Molecular Biology Institute. I would also like to thank my advisor, Prof. Tracy Johnson. I extend full gratitude to my PhD committee members, Prof. Greg Payne, Prof. Mike Carey, Prof. Albert Courey and Prof. Jeff Long, for providing academic support, allyship and authentic advocacy during this dissertation experience. I also would like to thank the members of the Johnson lab for the fruitful scientific discussions and feedback over the years. Specifically, I would like to thank Dr. Stephen Douglass for his technical contributions and assistance as my project transitioned from bench-top to bioinformatics. I would also like to extend full gratitude to my previous summer research mentees, Frank Gutierrez and Olayemi Oladapo, for trusting me with their mentorship in bioinformatics during Summer 2017.

Chapter 3 contains analysis of previously published datasets made publically available through the Gene Expression Omnibus by the following publication: Bruzzone, M. J., Grunberg, S., Kubik, S., Zentner, G. E., & Shore, D. (2018). Distinct patterns of histone acetyltransferase and Mediator

deployment at yeast protein-coding genes. Genes Dev, 32(17-18), 1252-1265. doi:10.1101/gad.312173.118

The science will always be there, but without true compassion, mentorship and community, the scientist will not. I would like to thank my mentors whose support, encouragement and wisdom is in large part why I made it this far. I would like to thank the incomparable vessel of light, Dr. Diana Azurdia, whose heart-driven and steadfast mentorship never waned since the day I interviewed at UCLA. I am fully grateful for her wisdom as well as our bond and sisterhood. I would also like to thank Dr. Daniel Hart who was my first research mentor at the UCSF Cardiovascular Research Institute. Dr. Hart's enthusiasm, patience, empathy and stellar scientific training unveiled a reservoir of scientific excellence and potential I didn't know I had. Thank you, Dan. I also would like to thank Dr. Frank Bayliss of the San Francisco State University (SFSU) NIH-MARC program and Dr. Steve Weissman of the City College of San Francisco LSAMP-Bridge to the Baccalaureate program. I would also like to thank Mrs. Valarie Watson, the Founder and Director of the Alafia Dance Ensemble of San Francisco as well as Dr. Angela Armendariz, my previous supervisor at the Exploratorium of San Francisco.

Last and most certainly not least, I would like to thank God. I would also like to thank my extended kin and the community that I've grown with over the years--in and outside of UCLA. This dissertation would absolutely not be possible without the divine protection, genuine friends, family, supporters and larger community from whom have poured into me on this journey.

To get to this point in my life was by no means a solitary feat. I am eternally grateful for the overflowing abundance of community that has always surrounded me in times of triumph and in times of trial on this journey to my PhD. I was guided to be in community with an endlessly rich collective of soul friends and family and I take deep pride and care for this responsibility. Over the years, my ancestral, spiritual and physical community has been and continues to be my power and has been my life-source to enrich the generational legacies that have been left for me as well as the micro- and macro-legacies that I have so delicately and intentionally crafted over the past few years. It is abundantly clear that I have been called to this path for reasons much larger and wider than what this dissertation could possibly articulate. To this end, I echo the poignant words of Caribbean writer and activist, Audre Lorde:

> *"It is better to speak. Remembering that we were never meant to survive"*

VITA

Shawntel Okonkwo

EDUCATION

San Francisco State University	B.S. Biology: Concentration in Physiology	2014
City College of San Francisco	Biology Track	2012
Fresno City College	Biology Track	2011

RESEARCH EXPERIENCE

PhD Candidate under Dr. Tracy Johnson October 2014 – November, 2019

University of California, Los Angeles

Dissertation: Exploring the mechanistic and regulatory properties of Gcn5-dependent acetylation on pre-mRNA splicing in *Saccharomyces cerevisiae*.

Undergraduate Student Research Fellow under Dr. Daniel Hart June, 2012 – June, 2014

Cardiovascular Research Institute - University of California, San Francisco

Identified a novel role of general core transcription co-factor TATA binding protein associated factor 3 (TAF3) in human erythropoiesis--particularly in regards to the control of the ratio between hemoglobin and cell volume and in the dynamics of red blood cell maturation.

PUBLICATIONS

Pistis G, Okonkwo SU, Traglia M, Sala C, Shin S-Y, et al. (2013) Genome Wide Association Analysis of a Founder Population Identified TAF3 as a Gene for MCHC in Humans. PLoS ONE 8(7): e69206.

AWARDS/FELLOWSHIPS/INTERNSHIPS

Awards:

National Science Foundation GRIP Fellow Award with the Smithsonian Institution's National Museum of African American History and Culture (NMAAHC) 2018

UCLA Graduate Programs in Sciences Representative for American Association for the Advancement of Science (AAAS) Catalyzing Advocacy in Science and Engineering Workshop on Capitol Hill 2018

American Heart Association's Idea Hack-a-thon Team Winner 2018

American Heart Association

McKinsey & Co APD Diversity Impact Award Finalist 2018

McKinsey & Company

Travel Scholarship Award 2013

Society for Advancement of Chicanos and Native Americans in Science (SACNAS)

Biosciences Opportunities Preview Program Travel Scholarship
University of Wisconsin - Madison 2013

1st Place - Poster Presentation on "A Role of TATA Binding Protein Associated Factor 3 in Human Erythropoiesis"
Personalized Medicine 6.0: Targeted Therapeutics in Next Generation Sequencing Conference 2013

3rd Place in Overall Life Sciences - Poster Presentation on "A Role of TATA Binding Protein Associated Factor 3 in Human Erythropoiesis"
College of Science and Engineering Student Research Competition - San Francisco State University 2013

3rd Place in Overall Speech - Intramural Speech Competition Championship
Fresno City College 2010

2nd Place in Persuasive Speaking - Intramural Speech Competition
Fresno City College 2009

Fellowships:

Kenneth I. Shine Fellowship 2017 – 2019

Bioscience Fellowship Incentive Program 2016 – 2019

NSF Graduate Research Fellowship Program *(Awarded)* 2016 – 2019

NSF Louis Stokes Alliance for Minority Participation (LSAMP) Bridge to the Doctorate Fellowship *(Awarded)* 2014 – 2020

Eugene V. Cota Robles Fellowship 2014 – 2020

National Research Council of the National Academies of Science, Engineering and Mathematics - Ford Pre-Doctoral Fellowship *(Offered and Declined)* 2016

National Research Council of the National Academies - Ford Pre-Doctoral Fellowship *(Honorable Mention)* 2015

UCLA Competitive Edge: NSF Summer Transition Program to the Doctorate Fellowship 2014

NIH Minority Access to Research Careers (MARC) Honors Fellowship Program 2012 –2014

NSF California State University Louis Stokes Alliance for Minority Participation (CSU-LSAMP) 2012 –2014

Internships:

NIH CCSF/SFSU Bridges to Baccalaureate Independent Research Internship 2012 - 2012

CCSF-Exploratorium Biology Laboratory Internship 2011 - 2011

RELEVANT SPEAKING ENGAGEMENT:

Okonkwo, SU. (2019) Intersectionality will save the future of science — Invited TEDx Speaker
TEDxUCLA — Los Angeles, CA

CHAPTER 1 - Introduction

1.1 Introduction to gene regulation

Biological definitions of life often rest upon the understanding that an organism arose from a set of instructions encoded in molecules and the biochemical reactions that drive them. Deoxyribonucleic Acid (or DNA for short), is the molecule that sets the biological foundation for not only what initiates life, but allows it to persist and thrive. On a macro-scale, it is accepted that our interstellar universe is predicated on entropic reactions and there is a level of regulation that produces the meta-level balance we understand to be today. On a micro-scale, the parallel argument remains as our molecular universe is overwhelmingly complex, diverse, dynamic and requires regulation for balance and harmonious production.

The instructions present in DNA molecules are called genes. Genes are the units of DNA molecules that provide the instructions for life. The central dogma of molecular biology posits that the genes present in the DNA starting material are replicated for effective scale (DNA replication), then are transcribed to produce the necessary working material for carrying out these instructions (RNA transcription) followed by the penultimate step of producing the protein workhorses that sustain cellular activity (protein translation). Although seemingly straightforward, since inception, this dogmatic view of biology has been expanded to include the wide-reaching nuances that further define and elucidate how life works at this molecular level. There are also spatio-temporal contexts to consider when discussing the implications of the central dogma of molecular biology. Developmental biology over time, homeostasis and the balance for sustained life and health of the cell/organism that can productively respond to internal/external stress, changes in environment, aging and any other reactionary/proactive

needs. Additionally keeping in mind the sheer magnitude of the tasks required to give rise to and maintain a living organism for the duration of a lifetime, one must consider the amount of entropy, randomness and chaos that can potentially exist as a result. This in particular is one of the primary reasons why gene regulation emerged as a field—to study how and why organisms are able to, in an organized and systematic manner, sustain life despite a molecular genetic landscape that is complex, deeply interconnected and primed for chaos and disorder.

Over the past few decades, our understanding of the regulation of gene expression has evolved. Until recent decades, the linear understanding of the central dogma of molecular biology drove sentiments around genes either being activated or repressed. Activation of genes was understood to be the functional precursor to the final gene product (the associated protein's structure and function in the cell) and for repression of genes, vice versa. We now understand that this seemingly unilateral biological process is highly ordered and involves an intricate series of events that give rise to complex outcomes at the extra-/intra-cellular and molecular level. There are numerous steps in gene regulation and many of these modes often participate in highly orchestrated cross-regulation--such as RNA transcription and processing.

1.2 Introduction to RNA transcription and co-transcriptional RNA processing

The immediately functional product of DNA is the ribonucleic acid (RNA) molecule. Although the main structural difference between DNA and RNA is the presence of a hydroxyl group on the 2' carbon instead of a hydrogen group, the functional differences vary considerably. RNA is produced by biochemical reactions orchestrated by an RNA polymerase enzyme with DNA as

the template and reaction substrate. In eukaryotic organisms, the RNA Polymerase II (RNAPII) holoenzyme is the protein responsible for transcribing messenger RNA from the DNA template. Through a series of reactions, RNAPII forms step-dependent complexes with other modular proteins throughout the transcription reaction in order to produce nascent mRNA (or pre-mRNA).

Pre-mRNA is the native product of RNAPII transcription from the DNA template and as such, this nascent molecule must undergo a considerable amount of processing and quality control to become mature messenger RNA. This processing is necessary for the messenger RNA to encode the proper functional instructions for the destined proteins as well as survive the vulnerable transport from the nucleus to the cytoplasmic ribosomes for protein translation. Pre-mRNA processing can occur co-transcriptionally--that is simultaneously with the RNA molecule being synthesized from RNAPII. For example, the 5' end of the nascent RNA molecule (specifically, the exposed guanosine) is methylated which produces a protective cap structure around the exposed end of the RNA. This 5' 7-methylguanosine cap not only protects the 5' end of the pre-mRNA molecule from degradation by exonucleases, but also functions to promote protein translation and regulate nuclear export. The most fascinating aspect of this molecular process is that this regulatory process occurs co-transcriptionally--functioning to court and prepare the RNA molecule for other modes of gene processing and regulation. In yeast, the enzymes associated with 5' capping of pre-mRNA include guanylyltransferase Ceg1p, methyltransferase Abd1p and triphosphatase Cet1p. These factors are recruited to RNAPII and deployed in synergy with the C-terminal domain (CTD) phosphorylation states of RNAPII--the transition from

transcriptional initiation to elongation--in both yeast and humans (Cho et al., 1998; Ho et al., 1998; Ho and Shuman, 1999). The regulated association and dissociation of these RNA processing factors is in direct synergy with the phosphorylation and dephosphorylation states of RNAPII CTD during transcription (Komarnitsky et al., 2000; Schroeder et al., 2000). This is just one example of how RNA processing is implicated during gene regulation.

Co-transcriptional gene regulation often occurs in a spatio-temporal manner. The coupling of RNAPII transcription and RNA processing events does not occur solely for the compartmentalization of molecular time. There is a large body of research that investigates the coupling of transcription and pre-mRNA processing. Hypotheses range from general cellular resource allocation to high-level quality control to other explanations regarding the economics of the cell. While pre-mRNA processing hosts a number of mechanisms for regulation in order to give rise to a mature mRNA product primed for protein translation, the primary mode of regulation explored here is pre-mRNA splicing.

1.3 The splicing of pre-mRNA is a key mRNA processing and quality control mechanism in eukaryotic organisms

The immediate product of RNAPII transcription is a nascent RNA molecule based on the preceding complementary DNA template, which as a result, contains both protein-coding RNA sequences (exons) and non-protein-coding sequences (introns). In order for these precursor RNA molecules to give rise to functional protein products, they must be further processed into mature mRNA molecules and made amenable for ribosomal translation in the cell. This key

quality-control step in mRNA processing is known as pre-mRNA splicing which occurs in the nucleus. Introns are removed from nascent RNA molecules and the resulting exonic sequences are spliced together, giving rise to mature mRNA (Berget et al., 1977; Chow et al., 1977).

The splicing reaction of pre-mRNA occurs through an array of coordinated biochemical mechanisms involving the step-wise assembly of the spliceosome. The spliceosome is a large ribonucleoprotein complex consisting of over 100 splicing factors and small nuclear ribonucleoproteins (snRNPs) which dynamically interact and rearrange over the splicing reaction (Brow, 2002; Matlin and Moore, 2007; Staley and Woolford, 2009). In yeast, each intron-containing pre-mRNA transcript contains highly-conserved sequences that are specifically recognized by components of the spliceosome during various steps of the splicing reaction for sequence binding, catalysis and exon ligation--namely, the donor site (5' splice site; 5' SS), the branchpoint sequence (BPS) and the acceptor (3' splice site; 3' SS) (Burge et al., 1999). In yeast, the typical consensus for both 5' SS and 3' SS are defined by dinucleotides 'GU' and 'AG', respectively. However, the branchpoint sequence tends to be defined by a longer 7-nucleotide notation--UACUAAC. Mutation of and/or lack of consensus in these splice site sequences may slightly disrupt or completely inhibit the assembly of the spliceosome onto the transcript and therefore the efficiency of the splicing reaction (Weiner and Zhuang, 1986; Wimmer et al., 2007).

Although a highly dynamic process involving over 100 proteins and structural rearrangements, splicing is generally driven by two transesterification reactions on the pre-mRNA substrate

(Moore et al., 1993). During the first reaction, the bulged 2’ -OH group on the BPS adenosine nucleotide performs a nucleophilic attack on the 5’ SS, resulting in intron cleavage (from the 5’ end of the transcript) and formation of an intermediate lariat structure (of the intron and 3’ end of the transcript). In the second reaction of splicing, the free 3’ -OH group of the 5’ exon attacks the 3’ ss to finally ligate the two exon products together.

One important aspect of pre-mRNA splicing is its role in establishing proper protein diversity in the organism. In higher eukaryotes such as mammals, most RNA transcripts are alternatively spliced to give rise to protein diversity that will subsequently meet a variety of needs for the cell.

1.4 Chromatin is a structural and functional regulatory platform for gene regulation

In Eukaryotes, DNA contains the critical genetic instructions for initiating and maintaining life in a living organism. Depending on the Eukaryotic organism, there can be anywhere from hundreds to tens of thousands of genes across the genome, and these genes undergo complex spatio-temporal conformations in order to regulate gene expression at any given time in development. In light of the high order of complexity and spatial constraints to gene regulation at the molecular level, physical structure is another dimension by which to regulate gene expression.

Fortunately, eukaryotes have evolved to accommodate the growing complexity of the three-dimensional genome and all of its dynamicity through space, time and environmental

stimuli. Chromatin is the physical scaffold for nuclear DNA to compress upon itself and is the structural combination of DNA, RNA and proteins. Throughout the eukaryotic genome, lengths of DNA molecules are wrapped around the macromolecular nucleosome structures, which effectively condenses the amount of DNA per unit of space in the nucleus.

Nucleosomes, the fundamental unit of chromatin, wrap approximately 147bp of DNA and are highly repeated throughout the genome (Luger et al., 1997). This level of compaction reduces the structural burden of encasing enough DNA in the cell nucleus for effective gene expression, but also leaves a new problem of chromatin accessibility to gene expression machinery. Nucleosomes consist of an octamer of histones, which have expanded beyond straightforward structural implications (i.e. chromatin compaction) for gene regulation into functional implications.

Histone proteins contain unique structural features that provide a platform for additional gene regulation, specifically post-translational modifications at their N-termini. A wide array of nuclear enzymes add methyl, acetyl and other chemical groups to the amino acid residues of the N-terminal tails of these histones, which in turn biochemically configure the structure of chromatin. Two main conformational states of chromatin often referenced are euchromatin and heterochromatin--the former referring to loosely bound chromatin and the latter referring to tightly bound chromatin. Each physical state has functional implications for gene expression as the loose state of DNA wrapped around euchromatin affords accessibility to transcriptional machinery. In contrast, heterochromatin is too tightly condensed to allow gene activation to

occur. Certain post-translational modifications of the histone N-terminal tails facilitate these chromatin states. For example, histone acetylation is the addition of an acetyl functional group from bioavailable Acetyl-CoA to Lysine residues on histone tails. When this enzymatic reaction occurs, the positively charged histone tail is neutralized (due to the negative charge of the acetyl group). The interaction between negatively charged DNA and the positively charged histones is weakened by this acetylation which loosens the interaction of the DNA with the nucleosome. Histone acetylation is often associated with gene activation as this biochemical reaction produces euchromatin and thus, an accessible environment for transcription factors to physically bind to gene promoters and initiate transcription.

2.1 What is co-transcriptional splicing and what is its relevance to gene regulation?

RNA processing events such as splicing, 5' capping and polyadenylation have been shown to occur simultaneously in concert with the process of transcription whereby the eukaryotic C-terminal domain (CTD) of the Rpb1 subunit of RNA Polymerase II holoenzyme (RNAPII) can serve as a physical regulatory platform for these processes. The step-wise assembly of the spliceosome onto pre-mRNA occurs co-transcriptionally. In addition to ensuring proper constitutive splicing for all intron-containing genes, the co-transcriptional nature of the splicing machinery facilitates the production of alternative transcripts from the same RNA template. Although the co-transcriptional nature of splicing is emerging as a well-researched topic, the mechanisms underlying the coupling and coordination of these reactions have not been well elucidated. These studies examine the mechanisms by which the highly dynamic chromatin

marks that are placed on histones during transcription affect splicing. They address a looming key question: How are these two reactions linked in the cell?

The splicing of pre-mRNA is carried out by the spliceosome, a large multi-subunit ribonucleoprotein complex which catalytically facilitates intron removal and exon ligation (Figure 1). The spliceosome is a highly dynamic and intricate molecular machine that assembles onto pre-mRNA in a stepwise manner. In each step of the splicing reaction, a multitude of biochemical conformations and RNA-protein interactions facilitate the recognition of canonical sequences within the intron, catalysis of the intronic sequences and ligation of the exon.
In the first commitment step of splicing, the U1 snRNP combined with other splicing factors recognizes the dinucleotide 5' splice site sequence (GU), which is positioned at the first 2 nucleotides of intronic RNA (Ruby and Abelson, 1988; Seraphin and Rosbash, 1989; Du and Rosbash, 2001). Following this early commitment step, the pre-spliceosome complex is formed as U2 snRNP complexed with enzymatic splicing factors recognize and bind to the branchpoint sequence (BPS) of the pre-mRNA substrate and expose the branch-point adenosine nucleotide for nucleophilic attack (Parker et al., 1987). In subsequent steps, the U4/U6-U5 tri-snRNP, in association with other splicing factors, complete the assembly of the spliceosome and further carry out the formation of the lariat intermediate, intron excision and exon ligation.

Co-transcriptional splicing is described as the simultaneous assembly of the spliceosome onto pre-mRNA while RNA Polymerase II and associated factors transcribe said pre-mRNA molecule (Figure 1C). As the 5' end of the nascent RNA molecule emerges from the exit tunnel of RNAPII,

the 5'SS is immediately recognized by splicing factors and spliceosome assembly begins. Early experiments to directly decipher the nature of this functional relationship were performed by Chromatin Immunoprecipitation (ChIP), which assays for direct DNA-protein interactions in cell-extracts (Gornemann et al., 2005; Lacadie and Rosbash, 2005; Wetterberg et al., 2001). The experimental design of ChIP does not accommodate direct interaction of protein with RNA, however, given the proximity of splicing components to the DNA template associated with the RNA in question, sound observations can be derived (Kotovic et al., 2003). The genome-wide observations of splicing-dependent RNAPII pausing events at exons provided robust and nuanced evidence to support the co-transcriptional nature of splicing (Carrillo Oesterreich et al., 2010; Alexander et al., 2010).

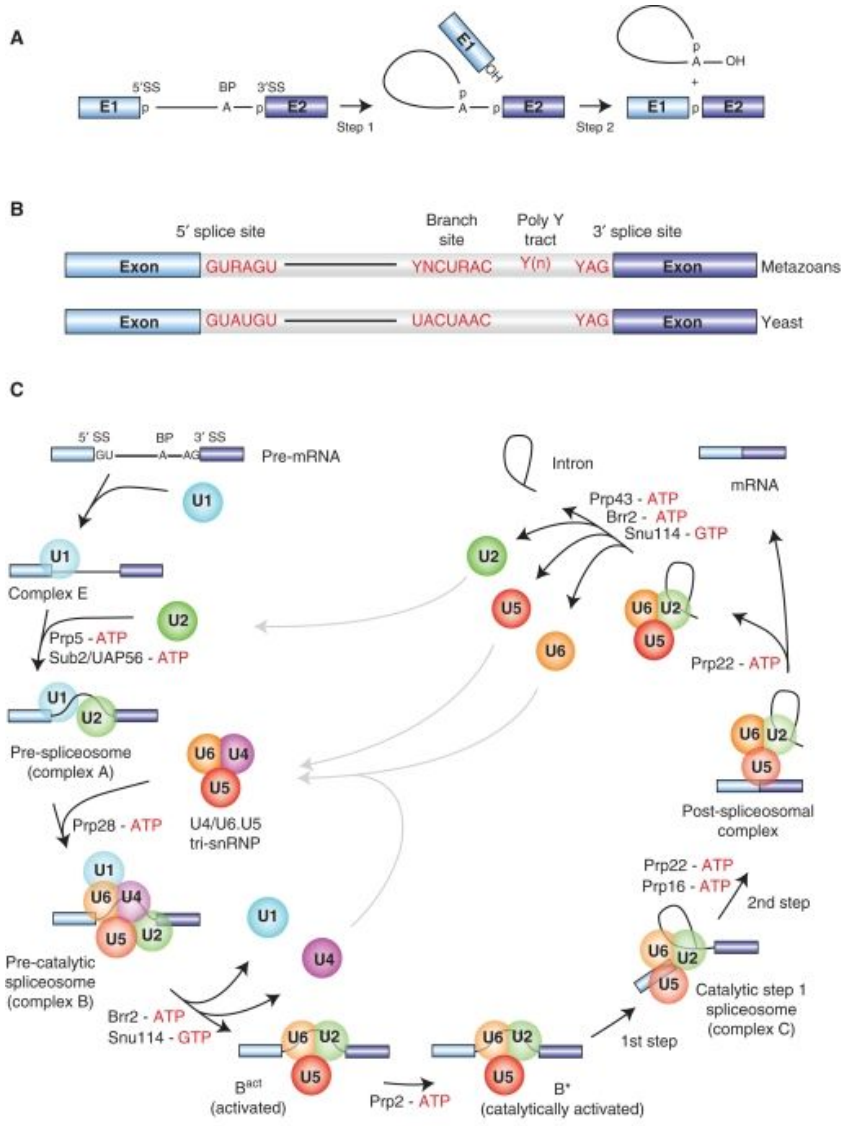


Figure 1. Pre-mRNA splicing and spliceosome assembly in yeast - A) The splicing of pre-mRNA in eukaryotes involves specific sequence-based motifs (5′ splice site (5′ SS), the 3′ SS and the branchpoint sequence (BPS). The splicing reaction occurs in two transesterification steps including lariat intermediate formation, catalysis and ligation of exons. Exons represented with blue and indigo boxes and introns with black line. B) The differences of splice site sequence motifs between yeast and metazoans. C) Spliceosome assembly contains a variety of RNA-protein interactions (Adapted from Will and Luhrmann, 2011).

Understanding the intersectional nature of co-transcriptional splicing with other modes of gene regulation at the mechanistic level is a field that is justly gaining attention as it begins to answer some of the fundamental questions in the splicing field: How and why is splicing linked to transcription? What evolutionary and biological purpose does this serve for the cell in a unicellular context as well as in the context of higher order organisms which have more sophisticated gene regulation programs and highly complex chromatin environments? In the context of chromatin, how does the highly dynamic nature of histone acetylation affect and correspond to the dynamic rearrangements of spliceosome assembly during transcription? Given the dynamic nature of co-transcriptional splicing, the composition of the spliceosome may also reflect the spatio-temporal priorities of the spliceosome in the co-transcriptional splicing reaction. For example, components of the transcriptional and chromatin environment have been shown to physically recruit components of the spliceosome to facilitate the splicing reaction (Leung et al., 2019; Luco et al., 2010).

2.2 Influence of chromatin on co-transcriptional splicing in yeast and mammals

Splicing is co-transcriptional, the transcription program is inherently linked to the chromatin environment and studies have shown that different components of the chromatin environment can interact with (and regulate) the spliceosome machinery. Given the broad range of chromatin modifications to explore, I will focus on histone acetylation and splicing. For example, depletion of a histone deacetylase increased H4 acetylation around alternative exonic junctions which lead to an increase in RNAPII elongation rate and subsequently decreased inclusion of the alternative exon. Gonzalez et al. demonstrated that lncRNA generated from the

human FGFR2 locus can regulate cell-specific alternative splicing through specific chromatin signatures (Gonzalez et al., 2015). A more direct relationship can be seen with Mrg15, a chromatin binding protein that specifically recognizes H3K36me3 marks and subsequently recruits the polypyrimidine-tract binding protein, Ptbp2, to the pre-mRNA substrate—facilitating exon-definition and thereby functioning as a direct regulatory link between the chromatin environment and the splicing machinery (Iwamori et al., 2016). Interestingly, single-gene studies on human genes *FN1* and *MCL1* have demonstrated a role for histone deacetylation by HDACs in regulating alternative splicing (Hnilicova et al., 2011; Khan et al., 2014; Choudhary et al., 2009; Gao and Koide, 2013). Finally, Schor and colleagues observed that perturbation of chromatin structure globally affects localization and recruitment of splicing factors (Schor et al., 2012).

A growing body of literature supports the role of nucleosomes in pre-mRNA splicing. One of the earliest observations of a relationship between splicing-relevant gene architecture and nucleosomes was led by Baldi et al., where they found naturally occurring nucleosome positioning signals in human exons and introns (Baldi et al., 1996). Here, they showed that exons (rather than introns) preferentially inhabit sequence triplets which have a thermodynamic preference for "locations on a bent double helix where the major groove faces inward and is compressed. The in-phase triplets are located adjacent to GCC/GGC triplets known to have the strongest bias in their positioning on the nucleosome." This observation was explored further by Denisov et al. as well as Kogan et al. as they rationalized the functional purpose for this sequence pattern to be a protection mechanism for splice sites (Denisov et al.,

1997; Kogan et al., 2005). Specifically, intronic splice sites on the template DNA strand are reported to be positioned near the center of the nucleosome particle for protection from UV-irradiation, nucleases and other cellular threats (Gale et al., 1987; Liu et al., 2000; Sollner-Webb et al., 1978; Thrall et al., 1994; Anderson and Widom, 2000). Lastly, an intriguing observation that single-nucleotide polymorphisms (SNPs) occur less frequently upstream of splice sites supports the protective model of nucleosomes to splice sites (Fairbrother et al., 2004).

With the emergence of sequencing technologies, there is now a substantial wealth of data by which research groups no longer need to explore each question from the bench. These innovations have inspired a new ethos of collaboration and multidisciplinarity in the scientific community, thereby reducing the pressure of time, increasing access to resources/facilities and shifting equity in intellectual pursuits between public and private institutions. To this concept, research groups over the years have begun analyzing publically available data-sets to answer new and hidden questions in a variety of contexts. This lends increased nuance to the variety of observations and questions that can be answered about chromatin and splicing in yeast. Additional observations on chromatin architecture and splicing have been uncovered through groups analyzing publically available data-sets derived by other members of the scientific community. Elegant work by Andersson et al. showed that there is a bias in nucleosomal positioning over exons as opposed to introns that is also independent of expression level in both humans and C. elegans (Andersson et al., 2009). Additionally, this group observed a higher incidence of histone modifications (largely H3K36me3, H3K79me1, H2BK5me1, H3K27me1,

H3K27me2, and H3K27me3) encompassing internal exons relative to the tandem intron (Andersson et al., 2009). From the Johnson lab, Leung and colleagues most recently uncovered a physical recruitment mechanism for the role of H3K36me3 in regulating constitutive splicing. This work showed that the chromodomain protein Eaf3 physically links the spliceosome to H3K36 methylation (Leung et al., 2019).

Through computational modeling of microarray data, Chen and colleagues observed higher occupancy of nucleosomes arranged at exonic sequences whereas intron sequences displayed lower occupancy in mammals (Chen et al., 2010). Additionally, by quantifying RNA structural flexibility with diversity of folding to energy ratios across transcripts, they observed that nucleosome occupied regions (exonic sequences) tend to have more rigid RNA structures while nucleosome-depleted regions had the opposite. While earlier groups have explored gene architecture within the context of nucleosomes, this data lends additional support for a relationship between nucleosome positioning and splice site definition.

Chromatin modifiers and signatures are now being understood to have a complex relationship with the splicing machinery. In the past 10 years, the Johnson lab has investigated multiple aspects of the relationship between chromatin and splicing. Since both of these features of gene regulation are individually complex, the relationships that synergize them are varied, nuanced and can be both specific and generalized. Our lab has previously shown that the major yeast histone acetyltransferase (HAT) GCN5, has genetic interactions with *MSL1* and *LEA1*—two U2 snRNP components of the spliceosome (Gunderson and Johnson, 2009). Additionally, we

observed that Gcn5 and its acetylation activity are required for proper co-transcriptional recruitment of the spliceosome in yeast (Gunderson et al., 2011). Specifically, U2 snRNP association with the branch-point of the pre-mRNA substrate and subsequent spliceosomal rearrangements are sensitive to Gcn5-dependent acetylation. Intriguingly, Gcn5-dependent acetylation is enriched at the promoter, but the HAT is found throughout the intron-containing genes, *DBP2* and *ECM33*. Deletion of histone deacetylases *HOS2* and *HOS3* reveals H3K9 and H3K14 acetylation throughout the gene bodies and abnormal persistence of U2 snRNP factors at the branch-point region of the pre-mRNA—with aberrant recruitment of downstream splicing factors (Gunderson et al., 2011).

The Johnson Lab demonstrated that Gcn5-dependent acetylation is important for spliceosomal rearrangements, however, it is not clear what the mechanism underlying this result is. The orchestration between HAT and HDAC activity on the DNA may be important for recruitment of factors that bind to chromatin and recruit the spliceosome or for tuning transcription elongation rates for proper stepwise recruitment of the spliceosome on the pre-mRNA substrate. Here, my thesis seeks to discover mechanistic insights into how Gcn5-mediated histone acetylation affects the splicing reaction genome-wide. Here, we seek to discover mechanistic insights into how Gcn5-mediated histone acetylation affects the splicing reaction.

2.3 Exploring co-transcriptional splicing in *Saccharomyces cerevisiae* as an ideal model organism

Addressing questions of co-transcriptional splicing in the model organism, *Saccharomyces cerevisiae* is ideal as it presents several advantages relevant to this research. Co-transcriptional splicing is a complex topic to experimentally investigate as it must account for the nuanced aspects and intrinsically dynamic features of transcription elongation, spliceosome assembly and the chromatin environment. Budding yeast is an organism that allows for an array of genetic and biochemical experimental manipulations to elegantly probe these mechanisms both in vivo and in vitro. For example, the genetic system in budding yeast is highly tractable, which provides the opportunity to create and combine synthetic mutations to test these questions. Additionally, the spliceosome machinery in the budding yeast is highly conserved between higher eukaryotes which ensures relevancy. Lastly, unlike the mammalian genome where variability of intron length, intron number and alternative splicing events for a single gene vary considerably, the genome of the budding yeast is more streamlined as about 5% of all genes contain introns (with a vast majority of these genes containing single introns which adhere to strong splice site consensus) and alternative splicing is limited.

3.1 The role of Gcn5 in histone acetylation and ribosomal gene regulation

In eukaryotes, activation of gene expression is regulated by a number of factors that are sequential, dynamic and highly regulated. As noted above, post-translational acetylation of lysine residues within the N-terminal tail of histone H3 is crucial to gene activation. This reaction is facilitated by the catalytic activity of Gcn5, the major yeast histone acetyltransferase, and Acetyl-CoA. Gcn5 recognizes and binds to lysine residues, catalyzes the transfer of an acetyl group from acetyl-CoA to the epsilon-amino group of lysines and activates

gene expression through histone-associated interactions with DNA (Tanner et al., 2000). The mechanism of action for this enzyme "utilizes a conserved glutamate residue to function as a general base through a bi-bi ternary complex mechanism (Trievel et al., 1999; Rojas et al., 1999; Clements et al., 1999). In this mechanism, both the acetyl-CoA and the substrate protein must be bound to the enzymes before catalysis can occur, and the turnover number for Gcn5 (kcat) is very efficient at ~210 min−1" (Poux et al., 2002).

While in a hetero-chromatinized (or closed) state, chromatin fibers are tightly wound--making it difficult for the pre-initiation complex of RNA-polymerase II to access the DNA for transcriptional activation with this type of locus often described as a repressed state, as described previously. Upon post-translational modification of the residues in proximity to the native DNA, the interaction between the histone proteins and the DNA weaken, thus allowing for the pre-initiation complex to assemble at the euchromatin locus and activate gene transcription.

Gcn5, a member of the P/CAF family of histone acetyltransferases, is known to acetylate most lysines on the N-terminus of histone H3, most notably being Lysine 9, 14, 18, 23 (Yang et al., 1996). *In vivo*, minor acetylation by Gcn5 has been observed on residues 8 and 16 of histone H4 and lysine residues in the amino-terminus of histone H2B (Grant et al., 1999; Zhang et al., 1998). It is suggested that due to the N-/C-terminal α2 and α4 helix residues interacting with the backbone of substrates rather than the side chains, Gcn5's fidelity for both histone and non-histone targets is high (Friedmann & Marmorstein, 2013). One of the major non-histone

targets of Gcn5 is the transactivator complex regulating ribosomal protein gene (RPG) expression.

Other than its role in post-translational histone modification, a number of reports have implicated Gcn5 in the regulation of RPG expression. For example, Gcn5 acetylates Ifh1, one of the master regulators of RPG expression--which effectively titrates it away from RPG promoters and thus aids in transcriptional repression of RPGs during nutrient stress (Downey et al., 2013). Additionally, acetylation by Gcn5 at the RPG cis-regulatory element *URS1*, is essential for activation of *IME2*, Ime1 and RSC, all critical components of the RPG expression program (Burgess et al., 2009).

4.1 Gcn5 plays multiple roles in cancer progression, maintenance, gene regulation and gene therapy

Being that Gcn5 is a crucial component of gene activation and chromatin accessibility, it plays a number of roles in maintaining cellular and genomic health/stability. Most of the abrogations in human health which implicate Gcn5 often associate with the role of the enzyme through indirect gene regulation and more direct protein acetylation. Interestingly, it was revealed that Gcn5 facilitates histone acetylation, nucleosomal eviction and RNA Polymerase II progression particularly at stress genes under stressed conditions in *S. pombe* (Sanso et al., 2011).

For example, many groups have reported the role of Gcn5 in cancer--this is separate from the additional roles in other non-cancerous human diseases. With c-Myc being heavily

over-expressed in cancer genetics, Gcn5 has been found to target this as a catalytic substrate as it stabilizes the protein through acetylation of lysine 323 on c-Myc itself (Flinn et al., 2002 & Patel et al., 2004). Additionally, Gcn5 has been found halt cellular differentiation in acute lymphoblastic leukemia (ALL) by acetylating and stabilizing the oncoprotein E2A-PBX1 and targeting HOX--an important factor in human cellular development (Holmlund et al., 2013 & Lu & Kamps, 1997). Gcn5 has also been implicated in breast cancer stem cell progression through other crucial human cellular development pathways such as the Wnt/beta-Catenin pathway (Chen et al., 2010). Interestingly, Gcn5 also acetylates the well-known tumor suppressor p53 to promote activation of its target genes (Barlev et al., 2001). Within breast cancer cells, Gcn5 more directly acetylates microtubule proteins via oncoproteins which serves a role in facilitating cancer cell movement (Li et al., 2015).

Gcn5 plays a major role in regulating genes associated with cancer via its histone acetylation activity. In 2016, researchers found that Gcn5 functions in promoting Hepatocellular Carcinoma (or liver cancer) progression by regulating and enhancing the expression of AIB1 (Amplified in Breast Cancer 1)--an oncogene strongly implicated in liver cancer (Majaz et al., 2016). Another role of Gcn5 regulating the gene expression in human disease has been uncovered in acetylating H3K9 on E2F1's promoter which enhances the expression of cell cycle control factors Cyclin D1 and Cyclin E--specifically promoting lung cancer progression. Not surprisingly, Gcn5 has also proven to be an attractive target for drug inhibitors, as inhibition of its HAT activity demonstrated effectiveness in blocking neuroblastoma cell growth (Gajer et al., 2015).

For decades, budding yeast has been used as a model for studying gene expression. Many of these studies in yeast aimed at understanding basic biology have uncovered and set the foundation for understanding mechanistic aspects of human disease. Through the strategic advantage of yeast genetics as a proxy for understanding human disease, many groups have successfully uncovered multiple roles for Gcn5 in human disease.

5. Genome-wide approaches for studying gene regulation

Until about 15 years ago, microarray analysis was the primary method for characterizing effects of gene mutation on the expression of a majority of genes. With the advent of next-generation sequencing, thoroughly answering sophisticated questions about gene regulation with the fine, base-pair resolution was finally possible. By coupling chemical manipulations (via crosslinking, nuclease and other chemical methods) to genome-wide surveys, multiple methods for understanding the full breadth of genomic features in a variety of mutant backgrounds and conditions was attainable. From exploring how the lack of a histone acetyltransferase affects gene expression across the genome through metabolic cycles to uncovering novel roles for otherwise canonical protein factors--genome-wide sequencing technologies afford deep excavation into the structural and functional mysteries of the genome. These methods will be explored in more detail as it pertains to Gcn5 in the following sections.

5.1 Types of mutations that elucidate Gcn5-dependent gene regulation genome-wide

gcn5Δ means that *GCN5* is genetically mutated and the Gcn5 protein is absent from the cell. While gene deletion is a straightforward and classic approach in understanding the functional

role of the gene product, gene deletion can lead to confounding results. Gcn5 is present and central to a number of cellular processes--such as histone acetylation, ribosomal protein gene regulation, pluripotency and cellular reprogramming as well as stress response. As the loss-of-function phenotype is viable, complete lack of Gcn5 in a cell can potentially trigger compensatory mechanisms for maintaining cellular stability.

The most conserved portion of Gcn5 is the HAT domain--the bromodomain region of the protein that interacts with and acetylates histone substrates (Brownell et al., 1996). This catalytic HAT domain spans amino acid residues 99 to 260 (Candau et al., 1997). It has also been long established that the substitution mutation, Y413A, in Gcn5 present in the ADA subcomplex of SAGA abgrogates acetyl-binding but does not affect the bromodomain or ADA complex structure (Syntichaki et al., 2000; Dhalluin et al., 1999; Li and Shogren-Knaak, 2009; Mujtaba et al., 2002).

Mutation of the HAT domain of Gcn5 is still in complex with SAGA, but catalytically inactive to bind and acetylate histone substrates. This type of mutation is typically utilized when aiming to understand the contribution of SAGA in full complex when Gcn5 is unable to acetylate histone or non-histone targets. Questions asked of this type of mutation could be whether Gcn5 can physically associate with specific target substrates outside of its catalytic role in histone acetylation.

Lastly, Gcn5 is known to interact with Ada2 and Ada5 within SAGA and the ADA subcomplex (Candau et al., 1997). This interaction facilitates nucleosomal acetylation as Gcn5 on its own (out of complex) confers weak acetylation to histone lysine substrates and is only strengthened in complex with SAGA as well as in the ADA subcomplex (Sterner et al., 2002; Candau et al., 1997). Deletion of the Ada2-interacting domain of Gcn5 (Gcn5 residues 260–280; just downstream of the HAT domain) results in an inability to target and acetylate nucleosomal histones (Candau et al., 1997). It has also been observed through structural studies that the SANT domain of Ada2 physically limits the Gcn5 Lys223 side chain from interacting with the Coenzyme A phosphate group (Sun et al., 2018). By virtue of these interactions, total deletion of Gcn5 or the Ada2-interacting domain will affect total histone acetyltransferase activity in the cell.

5.2 RNA-seq

There are many conclusions and observations one can infer about the genome from exploring the genomic DNA alone, however, transcriptome analysis offers a higher-resolution snapshot into the function of the genome. With the power of employing RNA-seq, that is, the Next-Generation sequencing technology that lends to understand genomic RNA populations in the cell in a high-throughput manner, researchers are now equipped to ask deeper questions about the spatio-temporal consequences of gene regulation.

Definition

RNA-seq allows for one to identify and quantify the type and amount of RNA in cell at any given moment, genetic background or experimental condition. The most popular application of RNA-seq is to explore differential gene expression in a cell after mutation or experimental treatment--comparing Wildtype to Experimental Condition (mutant, treatment, time-point). This information is valuable in creating a more holistic understanding of the gene regulatory landscape as the cell is constantly responding to the changing dynamics of the internal or external cellular environment. With RNA-seq, researchers can ask questions about different species of RNA in the cell--total RNA, ribosomal RNA, small RNA, microRNA, nuclear RNA, etc. The protocol is straightforward as it generally consists of RNA isolation, RNA selection or depletion and cDNA synthesis. RNA-seq can also be employed to understand RNA splicing, as exon, intron and exon-intron boundaries can be identified and analyzed in RNA-seq datasets.

Gcn5 relevant applications

There are many studies exploring the role of Gcn5 in basic and biomedical biology through identification of gene targets, expression profiling and co-regulation. As Gcn5 is a well studied transcriptional co-activator, these studies span widely across organisms and vary deeply in the questions prioritized. Here, I will focus on a few examples of the relevant applications of RNA-seq in exploring the role of Gcn5 in gene regulation.

The advent of RNA-seq provided new avenues for investigating how Gcn5 can affect gene expression and gene regulation upon external stimuli, over time or via mutation. For example, expression profiling through high-throughput RNA-seq demonstrated that Gcn5 regulates FGF

signaling and activates selective Myc target genes during early embryoid body differentiation (Wang et al., 2017). Furthermore, RNA-seq has been employed to explore the role of *GCN5* in alternative splicing that contributes to cellular reprogramming and pluripotency in mice. Through an RNAi screen, Hirsch and colleagues uncovered a co-regulatory network between Gcn5 and Myc which positively affects expression of an essential alternative splicing regulatory network during somatic cell reprogramming (Hirsch et al., 2015).

Over the years, Gcn5 has been shown to have crucial yet specific functions in various cellular physiologies (stress response, ribosomal protein gene regulation, pluripotency, histone acetylation and non-histone protein acetylation). RNA-seq allows researchers to dig deeper into how these functions are reflected in differential gene expression and by using the appropriate mutants it is possible to identify targets that are directly or indirectly regulated by Gcn5.

Additionally, as Gcn5 is an enzymatic protein, further questions regarding how the catalytically active or inactive state of Gcn5 affects gene expression in categories of genes and across the genome. In tandem with differential gene expression analysis, RNA-seq of samples mutated with Gcn5 (be it through point mutation, loss-of-function, over-expression or dosage-dependent mutation), pre-mRNA splicing can be surveyed as well. This information provides a more holistic understanding of the role of Gcn5 in gene regulation.

5.3 Chromatin Immunoprecipitation Sequencing (ChIP-seq)

Chromatin Immunoprecipitation (ChIP) is the molecular biology technique whereby proteins are cross-linked to DNA and immunoprecipitated to identify these proteins' genetic targets as well as relative enrichment at particular genomic loci. Combined with Next-Generation Sequencing, ChIP-seq brings a birds eye view of the abundance of particular protein or chromatin mark across the genome. This tool is powerful as it can uncover genome-wide regulation of transcription factors, chromatin modifiers, histone modifications as well as general proteins in a spatio-temporal context in any genetic or biochemical background. For example, assay provided key insights and confirmation into the mechanisms that regulate gene expression throughout stages in development via the ENCODE (Encyclopedia of DNA Elements) project (The ENCode Project Consortium, 2004).

Definition

ChIP-seq follows the same protocol of classic ChIP experimental set-up as the enrichment of loci bound by the protein of interest is the measurable outcome. Standard ChIP involves crosslinking total DNA to protein by a formaldehyde treatment followed by glycine to quench the reaction and prevent over-crosslinking. The DNA-protein complexes are then subjected to sonication to shear the DNA components into fragmented lengths optimized for accurate detection of the complex. Magnetic beads attached to the antibody of interest are added to the mixture of protein-bound DNA complexes to isolate the protein of interest. After immunoprecipitation and various cycles of high-/low-salt washing, the only complexes that remain should contain the protein of interest. The cross-links are reversed and the DNA is purified--leaving the DNA fragments enriched for the protein of interest. In ChIP-seq, these

fragments are sequenced then mapped to the genome and quantified based on signal to input. The resulting data uncovers a base-by-base resolution of enrichment for a protein or chromatin-associated mark of interest across the genome assayed.

Gcn5 relevant applications

Given the physical role of Gcn5 in gene activation and histone acetylation, ChIP-seq experiments provide a fantastic opportunity to understand the physical relationship of Gcn5 to the genome in various genetic and biochemical contexts. Many groups have employed ChIP-seq in various studies of Gcn5 occupancy, however, I will specifically focus on two examples. In HeLA cells, Gong and colleagues found that Gcn5 (along with Usp22) is bound to intergenic and intronic regions across the genome, but with few in the promoter region. They also found that Gcn5 can function in a diverse array of metabolic physiological processes in HeLA cells per the Gene Ontology (GO) analysis of the ChIP-seq data. The Usp22 and Gcn5 binding sites had similar distributions across the whole genome. Further analysis showed that the majority of Usp22 binding sites were located within intronic and intergenic regions, and approximately 58% of the peaks were located in intergenic, 34% in introns, 5% in exons and 3% in the promoter. For Gcn5 51% of the binding sites were in intergenic, 40% in introns, 5% in exons and 4% in the promoter (Gong et al., 2018). This result may challenge previous observations that Gcn5 is predominantly found at the promoter regions of protein coding genes, however, given the transient and quick-turnover nature of Gcn5 in acetylating histone residues, this data may rather reflect nuances in the dynamics of binding rather than static profiles of binding at these genetic

sequences. Potentially, increased binding for Gcn5 at the intronic regions of genes can suggest a more direct role for Gcn5 regulation at the level of the intron.

As mentioned earlier, cooperation between Gcn5-HAT, the SWI/SNF complex and the chaperone Ydj1 was uncovered primarily through ChIP-seq analyses with a variety of mutants. ChIP of histone H3 across the genome revealed that in *GCN5* depleted cells, histone H3 eviction was more defective compared to the other mutants. Deletion of gcn5 results in impaired H3 eviction upstream of the promoter at select Sulfurometron Methyl (SM)-induced groups of genes in the yeast genome. Per the report, the H3 histone at the -1 nucleosome remains at the promoter of these stress-induced genes much longer than in WT cells. The results of these analyses supported that that SWI/SNF, Gcn5, and Ydj1 cooperatively evict promoter nucleosomes and that through this mechanism, the most highly expressed genes in the yeast genome are most greatly affected by loss of any of these factors.

Bruzzone et al. employed a large genome-wide analysis using multiple NGS techniques to identify patterns of histone acetylation and Mediator deployment across the yeast protein-coding genome. Instead of drug-induced stress response or genetic depletion, Bruzzone and colleagues used the Anchor Away technique to deplete nuclear Gcn5 and ChIP-seq to survey RNA polymerase II association across the genome. They found that this depletion leads to a modestly decreased recruitment of RNAPII to nearly all yeast genes, with a moderate bias towards SAGA-dominated genes (Bruzzone et al., 2018). With the context of earlier studies showing Gcn5 having a modest effect on gene expression (with the exception of stress

conditions), these analyses may suggest other more complex roles for Gcn5 functioning in gene regulation.

One of the primary strengths of ChIP-seq is the ability to map DNA-protein interactions across the genome in any genetic or biochemical background. This information opens a wealth of understanding into the complexities of Gcn5 across the genome--from structure to function. Mapping chromatin-associated factors (such as histone modifications) provides a more complete picture for where and how Gcn5 is targeting its histone acetyltransferase activity.

Limitations

One of the primary weaknesses of using ChIP-seq is its inability to directly address questions involving RNA-protein interactions--particularly regarding pre-mRNA. Although many indirect insights can be and have been gathered using ChIP-seq (Gunderson and Johnson, 2009; Gunderson et al., 2011), the argument in the field remains on the premise that ChIP-seq assays DNA-protein interactions directly and not RNA-protein interactions. The general assumption is that the since DNA region(s) of interest will directly produce the pre-mRNA in question upon transcription, the co-transcriptional nature of spliceosomal assembly kinetics should tightly mirror what proteins bind to the associated DNA. General limitations of ChIP-seq revolve around optimizing experimental conditions which can affect the reproducibility of experimental data. Method of sonication (manual or automated) can affect the efficiency and yield of sonicated chromatin which can underestimate the number of DNA-protein interactions quantified in downstream analyses. Specificity as well as fidelity of antibodies used can cloud

the true representation of interaction--i.e. using elongating versus initiating RNAPII antibodies to assess polII occupancy across genes. Additionally, limitations of ChIP-seq can be seen in the interpretation of data where protein association with a genomic sequence does not explicitly define a direct functional relationship at the locus. Due to the forced nature of protein crosslinking early in the experimental prep, an overrepresentation of proteins in the vicinity of the locus can be confounded with proteins actually functioning on the locus in question.

5.4 Micrococcal Nuclease Sequencing (MNase-seq)

WIth most of the genome being punctuated by the structural nucleosomal particles that precede or respond to gene regulatory dynamics, it has become increasingly important to understand the localization of these nucleosomes through experimental conditions. MNase-seq follows roughly similar principles as ChIP-seq as the micrococcal nuclease is used to digest DNA that is not associated with nucleosomes, regions of the genome that are more or less sensitive to the treatment are sequenced then mapped and quantified based on signal to input ratio. MNase is often used to explore nucleosome localization, stability, and genomic structure in response to a genetic mutation, biochemical treatment or environmental change.

Definition

Micrococcal nuclease is an endo-/exo-nuclease enzyme used to digest DNA that is not associated with nucleosomes. In an MNase-seq experiment, MNase is added to cells at a particular concentration to introduce single-stranded followed by double-stranded breaks in DNA that is not associated with nucleosomes, and the reaction is halted with EDTA.

Nucleosome-associated DNA is insensitive to MNase (Axel, 1975; Sulkowski and Laskowski, 1962) and as such, the exonuclease property of MNase continues to digest DNA molecules until it reaches a nucleosome that is protected by DNA. The resulting undigested DNA-nucleosome particles are then sequenced and mapped to the genome--providing key information about the location of nucleosomes.

Gcn5 relevant applications

Bruzzone and colleagues were interested in exploring patterns of histone acetyltransferase and Mediator association at protein coding genes in yeast and chose to employ various NGS techniques to uncover key insights. In addition to other NGS techniques, MNase-seq was used in this paper to further explore the role of Gcn5 in transcription. The group found that nuclear depletion by anchor-away of Gcn5 did not affect nucleosome occupancy when surveying the top 500 genes that showed a >30% reduction in RNAPII under Gcn5 depletion (Bruzzone et al., 2018).

Strengths

One of the primary strengths of MNase-seq is in its ability to map nucleosomes across the genome. Prior to this, proxies for nucleosome localization were often left to ChIP-seq on nucleosome-associated proteins. Another strength is its ability to cover a lengthy spectrum of regulatory protein locations and binding sites in the genome as nucleosomes function as a major platform for gene regulatory activity. With regards to Gcn5, this aids in understanding more nuanced roles for Gcn5 in nucleosomal gene regulation.

Limitations

While MNase-seq offers a unique look into the regulatory landscape of Gcn5 with regards to nucleosome positioning, the identity of the regulatory elements that may span the MNase-sensitive region remains unknown in these types of datasets. As a consequence, MNase-seq data must be paired with ChIP-seq data in order to make sense of the regulatory landscape--for example, coupling enrichment of differentially modified H3 provide insight to the nature of the nucleosome at the position denoted by MNase-seq. Additionally, older studies have noted that the sequence-specificity of MNase enzyme can introduce bias into MNase-derived datasets (Horz and Altenburger, 1981; Flick et al., 1986; Noll and Kornberg, 1977). Lastly, incomplete digestion by MNase in the wet-bench phase of the experimental prep can be an issue when quantifying results during the data analysis stage--under-estimating the true results.

6.1 Rationale for genome-wide exploration of Gcn5 in splicing

The Johnson lab has shown that Gcn5-dependent histone acetylation affects co-transcriptional spliceosome assembly and splicing efficiency of two intron-containing genes, however it was not known how the enzyme affects splicing on a genome-wide scale. A proper survey of the global co-transcriptional splicing landscape under depletion of Gcn5 and its major histone target is necessary. Understanding the role of Gcn5-dependent histone acetylation in co-transcriptional splicing across the genome can provide broader yet more refined implications for its function in integrating chromatin dynamics with pre-mRNA splicing. For

example, does Gcn5 affect splicing efficiency of all intron-containing genes in the same manner? Are there particular structural or functional categories of genes that are more sensitive to Gcn5 depletion? Next-generation based sequencing data can provide suggestive mechanistic information for Gcn5 in splicing that is contextualized by the dynamic nature of chromatin and co-transcriptional splicing in the cell. For example, are there other processes in the gene regulatory program that may influence how Gcn5 is affecting co-transcriptional splicing? Are highly expressed intron-containing genes affected differently than lowly expressed genes? The structural and functional identity of the genes that are categorically affected by these mutations can provide deeper and more critical insights into how Gcn5-dependent histone acetylation influences co-transcriptional splicing genome-wide and this thesis aims to explore that. In Chapter 2, I explore the genome-wide role of *GCN5* in pre-mRNA splicing through RNA-seq (a first for the field) and in Chapter 3, I provide evidence for the role of chromatin architecture determining Gcn5-dependent splicing outcomes as a putative mechanism. Lastly, in Chapter 4, I summarize the major findings of the data and offer future directions to future researchers directly or indirectly interested in investigating this topic collaboratively.

REFERENCES

Alexander, R. D., Innocente, S. A., Barrass, J. D., & Beggs, J. D. (2010). Splicing-dependent RNA polymerase pausing in yeast. *Mol Cell, 40*(4), 582-593. doi:10.1016/j.molcel.2010.11.005

Anderson, J. D., & Widom, J. (2000). Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *J Mol Biol, 296*(4), 979-987. doi:10.1006/jmbi.2000.3531

Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C., & Komorowski, J. (2009). Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res, 19*(10), 1732-1741. doi:10.1101/gr.092353.109

Axel, R. (1975). Cleavage of DNA in nuclei and chromatin with staphylococcal nuclease. *Biochemistry, 14*(13), 2921-2925. doi:10.1021/bi00684a020

Baldi, P., Brunak, S., Chauvin, Y., & Krogh, A. (1996). Naturally occurring nucleosome positioning signals in human exons and introns. *J Mol Biol, 263*(4), 503-510. doi:10.1006/jmbi.1996.0592

Barlev, N. A., Liu, L., Chehab, N. H., Mansfield, K., Harris, K. G., Halazonetis, T. D., & Berger, S. L. (2001). Acetylation of p53 activates transcription through recruitment of coactivators/histone acetyltransferases. *Mol Cell, 8*(6), 1243-1254. doi:10.1016/s1097-2765(01)00414-2

Berget, S. M., Moore, C., & Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A, 74*(8), 3171-3175. doi:10.1073/pnas.74.8.3171

Brow, D. A. (2002). Allosteric cascade of spliceosome activation. *Annu Rev Genet, 36*, 333-360. doi:10.1146/annurev.genet.36.043002.091635

Brownell, J. E., & Allis, C. D. (1996). Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Curr Opin Genet Dev, 6*(2), 176-184. doi:10.1016/s0959-437x(96)80048-7

Bruzzone, M. J., Grunberg, S., Kubik, S., Zentner, G. E., & Shore, D. (2018). Distinct patterns of histone acetyltransferase and Mediator deployment at yeast protein-coding genes. *Genes Dev, 32*(17-18), 1252-1265. doi:10.1101/gad.312173.118

Burge, C. B., Tuschl, T., and Sharp, P. (1999). Splicing of precursors to mRNAs by the spliceosomes. *Cold Spring Harbor Monograph Series*.

Burgess, R. J., Guy, M. P., & Zhang, Z. (2009). Fueling transcriptional silencing with Gas1. *Proc Natl Acad Sci U S A, 106*(27), 10879-10880. doi:10.1073/pnas.0905192106

Candau, R., Zhou, J. X., Allis, C. D., & Berger, S. L. (1997). Histone acetyltransferase activity and interaction with ADA2 are critical for GCN5 function in vivo. *EMBO J, 16*(3), 555-565. doi:10.1093/emboj/16.3.555

Carrillo Oesterreich, F., Preibisch, S., & Neugebauer, K. M. (2010). Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol Cell, 40*(4), 571-581. doi:10.1016/j.molcel.2010.11.004

Chen, J., Luo, Q., Yuan, Y., Huang, X., Cai, W., Li, C., . . . Li, B. (2010). Pygo2 associates with MLL2 histone methyltransferase and GCN5 histone acetyltransferase complexes to augment Wnt target gene expression and breast cancer stem-like cell expansion. *Mol Cell Biol, 30*(24), 5621-5635. doi:10.1128/MCB.00465-10

Chen, W., Luo, L., & Zhang, L. (2010). The organization of nucleosomes around splice sites. *Nucleic Acids Res, 38*(9), 2788-2798. doi:10.1093/nar/gkq007

Cho, E. J., Rodriguez, C. R., Takagi, T., & Buratowski, S. (1998). Allosteric interactions between capping enzyme subunits and the RNA polymerase II carboxy-terminal domain. *Genes Dev, 12*(22), 3482-3487. doi:10.1101/gad.12.22.3482

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., . . . Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science, 325*(5942), 834-840. doi:10.1126/science.1175371

Chow, L. T., Gelinas, R. E., Broker, T. R., & Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell, 12*(1), 1-8. doi:10.1016/0092-8674(77)90180-5

Clements, A., Rojas, J. R., Trievel, R. C., Wang, L., Berger, S. L., & Marmorstein, R. (1999). Crystal structure of the histone acetyltransferase domain of the human PCAF transcriptional regulator bound to coenzyme A. *EMBO J, 18*(13), 3521-3532. doi:10.1093/emboj/18.13.3521

Consortium, E. P. (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science, 306*(5696), 636-640. doi:10.1126/science.1105136

Denisov, D. A., Shpigelman, E. S., & Trifonov, E. N. (1997). Protective nucleosome centering at splice sites as suggested by sequence-directed mapping of the nucleosomes. *Gene, 205*(1-2), 145-149. doi:10.1016/s0378-1119(97)00406-x

Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K., & Zhou, M. M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. *Nature, 399*(6735), 491-496. doi:10.1038/20974

Downey, M., Knight, B., Vashisht, A. A., Seller, C. A., Wohlschlegel, J. A., Shore, D., & Toczyski, D. P. (2013). Gcn5 and sirtuins regulate acetylation of the ribosomal protein transcription factor Ifh1. *Curr Biol, 23*(17), 1638-1648. doi:10.1016/j.cub.2013.06.050

Du, H., & Rosbash, M. (2001). Yeast U1 snRNP-pre-mRNA complex formation without U1snRNA-pre-mRNA base pairing. *RNA, 7*(1), 133-142. doi:10.1017/s1355838201001844

Fairbrother, W. G., Holste, D., Burge, C. B., & Sharp, P. A. (2004). Single nucleotide polymorphism-based validation of exonic splicing enhancers. *PLoS Biol, 2*(9), E268. doi:10.1371/journal.pbio.0020268

Flick, J. T., Eissenberg, J. C., & Elgin, S. C. (1986). Micrococcal nuclease as a DNA structural probe: its recognition sequences, their genomic distribution and correlation with DNA structure determinants. *J Mol Biol, 190*(4), 619-633. doi:10.1016/0022-2836(86)90247-0

Flinn, E. M., Wallberg, A. E., Hermann, S., Grant, P. A., Workman, J. L., & Wright, A. P. (2002). Recruitment of Gcn5-containing complexes during c-Myc-dependent gene activation. Structure and function aspects. *J Biol Chem, 277*(26), 23399-23406. doi:10.1074/jbc.M201704200

Friedmann, D. R., & Marmorstein, R. (2013). Structure and mechanism of non-histone protein acetyltransferase enzymes. *FEBS J, 280*(22), 5570-5581. doi:10.1111/febs.12373

Gajer, J. M., Furdas, S. D., Grunder, A., Gothwal, M., Heinicke, U., Keller, K., . . . Jung, M. (2015). Histone acetyltransferase inhibitors block neuroblastoma cell growth in vivo. *Oncogenesis, 4*, e137. doi:10.1038/oncsis.2014.51

Gale, J. M., Nissen, K. A., & Smerdon, M. J. (1987). UV-induced formation of pyrimidine dimers in nucleosome core DNA is strongly modulated with a period of 10.3 bases. *Proc Natl Acad Sci U S A, 84*(19), 6644-6648. doi:10.1073/pnas.84.19.6644

Gao, Y., & Koide, K. (2013). Chemical perturbation of Mcl-1 pre-mRNA splicing to induce apoptosis in cancer cells. *ACS Chem Biol, 8*(5), 895-900. doi:10.1021/cb300602j

Gong, Z., Liu, J., Xie, X., Xu, X., Wu, P., Li, H., . . . Xiong, J. (2018). Identification of potential target genes of USP22 via ChIP-seq and RNA-seq analysis in HeLa cells. *Genet Mol Biol, 41*(2), 488-495. doi:10.1590/1678-4685-GMB-2017-0164

Gonzalez, I., Munita, R., Agirre, E., Dittmer, T. A., Gysling, K., Misteli, T., & Luco, R. F. (2015). A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. *Nat Struct Mol Biol, 22*(5), 370-376. doi:10.1038/nsmb.3005

Gornemann, J., Kotovic, K. M., Hujer, K., & Neugebauer, K. M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol Cell, 19*(1), 53-63. doi:10.1016/j.molcel.2005.05.007

Grant, P. A., Eberharter, A., John, S., Cook, R. G., Turner, B. M., & Workman, J. L. (1999). Expanded lysine acetylation specificity of Gcn5 in native complexes. *J Biol Chem, 274*(9), 5895-5900. doi:10.1074/jbc.274.9.5895

Gunderson, F. Q., & Johnson, T. L. (2009). Acetylation by the transcriptional coactivator Gcn5 plays a novel role in co-transcriptional spliceosome assembly. *PLoS Genet, 5*(10), e1000682. doi:10.1371/journal.pgen.1000682

Gunderson, F. Q., Merkhofer, E. C., & Johnson, T. L. (2011). Dynamic histone acetylation is critical for cotranscriptional spliceosome assembly and spliceosomal rearrangements. *Proc Natl Acad Sci U S A, 108*(5), 2004-2009. doi:10.1073/pnas.1011982108

Hirsch, C. L., Coban Akdemir, Z., Wang, L., Jayakumaran, G., Trcka, D., Weiss, A., . . . Dent, S. Y. (2015). Myc and SAGA rewire an alternative splicing network during early somatic cell reprogramming. *Genes Dev, 29*(8), 803-816. doi:10.1101/gad.255109.114

Hnilicova, J., Hozeifi, S., Duskova, E., Icha, J., Tomankova, T., & Stanek, D. (2011). Histone deacetylase activity modulates alternative splicing. *PLoS One, 6*(2), e16727. doi:10.1371/journal.pone.0016727

Ho, C. K., & Shuman, S. (1999). Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol Cell, 3*(3), 405-411. doi:10.1016/s1097-2765(00)80468-2

Ho, C. K., Sriskanda, V., McCracken, S., Bentley, D., Schwer, B., & Shuman, S. (1998). The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem, 273*(16), 9577-9585. doi:10.1074/jbc.273.16.9577

Holmlund, T., Lindberg, M. J., Grander, D., & Wallberg, A. E. (2013). GCN5 acetylates and regulates the stability of the oncoprotein E2A-PBX1 in acute lymphoblastic leukemia. *Leukemia, 27*(3), 578-585. doi:10.1038/leu.2012.265

Horz, W., & Altenburger, W. (1981). Sequence specific cleavage of DNA by micrococcal nuclease. *Nucleic Acids Res, 9*(12), 2643-2658. doi:10.1093/nar/9.12.2643

Iwamori, N., Tominaga, K., Sato, T., Riehle, K., Iwamori, T., Ohkawa, Y., . . . Matzuk, M. M. (2016). MRG15 is required for pre-mRNA splicing and spermatogenesis. *Proc Natl Acad Sci U S A, 113*(37), E5408-5415. doi:10.1073/pnas.1611995113

Khan, D. H., Gonzalez, C., Cooper, C., Sun, J. M., Chen, H. Y., Healy, S., . . . Davie, J. R. (2014). RNA-dependent dynamic histone acetylation regulates MCL1 alternative splicing. *Nucleic Acids Res, 42*(3), 1656-1670. doi:10.1093/nar/gkt1134

Kogan, S., & Trifonov, E. N. (2005). Gene splice sites correlate with nucleosome positions. *Gene, 352*, 57-62. doi:10.1016/j.gene.2005.03.004

Komarnitsky, P., Cho, E. J., & Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev, 14*(19), 2452-2460. doi:10.1101/gad.824700

Kotovic, K. M., Lockshon, D., Boric, L., & Neugebauer, K. M. (2003). Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. *Mol Cell Biol, 23*(16), 5768-5779. doi:10.1128/mcb.23.16.5768-5779.2003

Leung, C. S., Douglass, S. M., Morselli, M., Obusan, M. B., Pavlyukov, M. S., Pellegrini, M., & Johnson, T. L. (2019). H3K36 Methylation and the Chromodomain Protein Eaf3 Are Required for Proper Cotranscriptional Spliceosome Assembly. *Cell Rep, 27*(13), 3760-3769 e3764. doi:10.1016/j.celrep.2019.05.100

Li, L., Liu, B., Zhang, X., & Ye, L. (2015). The oncoprotein HBXIP promotes migration of breast cancer cells via GCN5-mediated microtubule acetylation. *Biochem Biophys Res Commun, 458*(3), 720-725. doi:10.1016/j.bbrc.2015.02.036

Li, S., & Shogren-Knaak, M. A. (2009). The Gcn5 bromodomain of the SAGA complex facilitates cooperative and cross-tail acetylation of nucleosomes. *J Biol Chem, 284*(14), 9411-9417. doi:10.1074/jbc.M809617200

Liu, X., Mann, D. B., Suquet, C., Springer, D. L., & Smerdon, M. J. (2000). Ultraviolet damage and nucleosome folding of the 5S ribosomal RNA gene. *Biochemistry, 39*(3), 557-566. doi:10.1021/bi991771m

Lu, Q., & Kamps, M. P. (1997). Heterodimerization of Hox proteins with Pbx1 and oncoprotein E2a-Pbx1 generates unique DNA-binding specificity at nucleotides predicted to contact the N-terminal arm of the Hox homeodomain--demonstration of Hox-dependent targeting of E2a-Pbx1 in vivo. *Oncogene, 14*(1), 75-83. doi:10.1038/sj.onc.1200799

Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M., & Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science, 327*(5968), 996-1000. doi:10.1126/science.1184208

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature, 389*(6648), 251-260. doi:10.1038/38444

Majaz, S., Tong, Z., Peng, K., Wang, W., Ren, W., Li, M., . . . Yu, C. (2016). Histone acetyl transferase GCN5 promotes human hepatocellular carcinoma progression by enhancing AIB1 expression. *Cell Biosci, 6*, 47. doi:10.1186/s13578-016-0114-6

Matlin, A. J., & Moore, M. J. (2007). Spliceosome assembly and composition. *Adv Exp Med Biol, 623*, 14-35. doi:10.1007/978-0-387-77374-2_2

Mizzen, C. A., Yang, X. J., Kokubo, T., Brownell, J. E., Bannister, A. J., Owen-Hughes, T., . . . Allis, C. D. (1996). The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell, 87*(7), 1261-1270. doi:10.1016/s0092-8674(00)81821-8

Moore, M. J., & Sharp, P. A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature, 365*(6444), 364-368. doi:10.1038/365364a0

Mujtaba, S., He, Y., Zeng, L., Farooq, A., Carlson, J. E., Ott, M., . . . Zhou, M. M. (2002). Structural basis of lysine-acetylated HIV-1 Tat recognition by PCAF bromodomain. *Mol Cell, 9*(3), 575-586. doi:10.1016/s1097-2765(02)00483-5

Noll, M., & Kornberg, R. D. (1977). Action of micrococcal nuclease on chromatin and the location of histone H1. *J Mol Biol, 109*(3), 393-404. doi:10.1016/s0022-2836(77)80019-3

Parker, R., Siliciano, P. G., & Guthrie, C. (1987). Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell, 49*(2), 229-239. doi:10.1016/0092-8674(87)90564-2

Patel, J. H., Du, Y., Ard, P. G., Phillips, C., Carella, B., Chen, C. J., . . . McMahon, S. B. (2004). The c-MYC oncoprotein is a substrate of the acetyltransferases hGCN5/PCAF and TIP60. *Mol Cell Biol, 24*(24), 10826-10834. doi:10.1128/MCB.24.24.10826-10834.2004

Poux, A. N., Cebrat, M., Kim, C. M., Cole, P. A., & Marmorstein, R. (2002). Structure of the GCN5 histone acetyltransferase bound to a bisubstrate inhibitor. *Proc Natl Acad Sci U S A, 99*(22), 14065-14070. doi:10.1073/pnas.222373899

Rojas, J. R., Trievel, R. C., Zhou, J., Mo, Y., Li, X., Berger, S. L., . . . Marmorstein, R. (1999). Structure of Tetrahymena GCN5 bound to coenzyme A and a histone H3 peptide. *Nature, 401*(6748), 93-98. doi:10.1038/43487

Ruby, S. W., & Abelson, J. (1988). An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. *Science, 242*(4881), 1028-1035. doi:10.1126/science.2973660

Sanso, M., Vargas-Perez, I., Quintales, L., Antequera, F., Ayte, J., & Hidalgo, E. (2011). Gcn5 facilitates Pol II progression, rather than recruitment to nucleosome-depleted stress promoters, in Schizosaccharomyces pombe. *Nucleic Acids Res, 39*(15), 6369-6379. doi:10.1093/nar/gkr255

Schor, I. E., Lleres, D., Risso, G. J., Pawellek, A., Ule, J., Lamond, A. I., & Kornblihtt, A. R. (2012). Perturbation of chromatin structure globally affects localization and recruitment of splicing factors. *PLoS One, 7*(11), e48084. doi:10.1371/journal.pone.0048084

Schroeder, S. C., Schwer, B., Shuman, S., & Bentley, D. (2000). Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev, 14*(19), 2435-2440. doi:10.1101/gad.836300

Seraphin, B., & Rosbash, M. (1989). Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell, 59*(2), 349-358. doi:10.1016/0092-8674(89)90296-1

Sollner-Webb, B., Melchior, W., Jr., & Felsenfeld, G. (1978). DNAase I, DNAase II and staphylococcal nuclease cut at different, yet symmetrically located, sites in the nucleosome core. *Cell, 14*(3), 611-627. doi:10.1016/0092-8674(78)90246-5

Staley, J. P., & Woolford, J. L., Jr. (2009). Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. *Curr Opin Cell Biol, 21*(1), 109-118. doi:10.1016/j.ceb.2009.01.003

Sterner, D. E., Wang, X., Bloom, M. H., Simon, G. M., & Berger, S. L. (2002). The SANT domain of Ada2 is required for normal acetylation of histones by the yeast SAGA complex. *J Biol Chem, 277*(10), 8178-8186. doi:10.1074/jbc.M108601200

Sulkowski, E., & Laskowski, M., Sr. (1962). Mechanism of action of micrococcal nuclease on deoxyribonucleic acid. *J Biol Chem, 237*, 2620-2625.

Sun, J., Paduch, M., Kim, S. A., Kramer, R. M., Barrios, A. F., Lu, V., . . . Tan, S. (2018). Structural basis for activation of SAGA histone acetyltransferase Gcn5 by partner subunit Ada2. *Proc Natl Acad Sci U S A, 115*(40), 10010-10015. doi:10.1073/pnas.1805343115

Syntichaki, P., Topalidou, I., & Thireos, G. (2000). The Gcn5 bromodomain co-ordinates nucleosome remodelling. *Nature, 404*(6776), 414-417. doi:10.1038/35006136

Tanner, K. G., Langer, M. R., Kim, Y., & Denu, J. M. (2000). Kinetic mechanism of the histone acetyltransferase GCN5 from yeast. *J Biol Chem, 275*(29), 22048-22055. doi:10.1074/jbc.M002893200

Thrall, B. D., Mann, D. B., Smerdon, M. J., & Springer, D. L. (1994). Nucleosome structure modulates benzo[a]pyrenediol epoxide adduct formation. *Biochemistry, 33*(8), 2210-2216. doi:10.1021/bi00174a030

Trievel, R. C., Rojas, J. R., Sterner, D. E., Venkataramani, R. N., Wang, L., Zhou, J., . . . Marmorstein, R. (1999). Crystal structure and mechanism of histone acetylation of the

yeast GCN5 transcriptional coactivator. *Proc Natl Acad Sci U S A, 96*(16), 8931-8936. doi:10.1073/pnas.96.16.8931

Wang, L., Scott, I., Zhu, L., Wu, K., Han, K., Chen, Y., . . . Sack, M. N. (2017). GCN5L1 modulates cross-talk between mitochondria and cell signaling to regulate FoxO1 stability and gluconeogenesis. *Nat Commun, 8*(1), 523. doi:10.1038/s41467-017-00521-8

Wetterberg, I., Zhao, J., Masich, S., Wieslander, L., & Skoglund, U. (2001). In situ transcription and splicing in the Balbiani ring 3 gene. *EMBO J, 20*(10), 2564-2574. doi:10.1093/emboj/20.10.2564

Will, C. L., & Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb Perspect Biol, 3*(7). doi:10.1101/cshperspect.a003707

Wimmer, K., Roca, X., Beiglbock, H., Callens, T., Etzler, J., Rao, A. R., . . . Messiaen, L. (2007). Extensive in silico analysis of NF1 splicing defects uncovers determinants for splicing outcome upon 5' splice-site disruption. *Hum Mutat, 28*(6), 599-612. doi:10.1002/humu.20493

Zhang, W., Bone, J. R., Edmondson, D. G., Turner, B. M., & Roth, S. Y. (1998). Essential and redundant functions of histone acetylation revealed by mutation of target lysines and loss of the Gcn5p acetyltransferase. *EMBO J, 17*(11), 3155-3167. doi:10.1093/emboj/17.11.3155

Zhuang, Y., & Weiner, A. M. (1986). A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell, 46*(6), 827-835. doi:10.1016/0092-8674(86)90064-4

CHAPTER 2 - Loss of Gcn5-mediated histone acetylation demonstrates categorical effects on regulation of gene expression and splicing genome-wide

2.1 Introduction

In Eukaryotes, pre-mRNA splicing occurs in a co-transcriptional manner. As RNA Polymerase II transcribes nascent RNA molecules, the spliceosome sequentially assembles onto these pre-mRNA transcripts simultaneously. There are vast implications for the chromatin environment playing a role in facilitating co-transcriptional splicing, as RNAPII is heavily influenced by dynamics in the chromatin environment. Events such as chromatin remodeling, nucleosome positioning and post-translational modification of histones have been shown to influence splicing in both yeast and mammals.

The step-wise assembly of the spliceosome onto pre-mRNA occurs co-transcriptionally, while the nascent transcript is synthesized from RNA Polymerase II. The mechanisms underlying the molecular coupling and coordination of these co-transcriptional reactions have not been thoroughly elucidated in the context of the dynamic chromatin environment. The Johnson Lab previously discovered that the major yeast histone acetyltransferase (HAT) Gcn5 demonstrates genetic interactions with *MSL1* and *LEA1*—two core U2 small nuclear ribonucleoprotein particle (snRNP) components of the spliceosome. Additionally, they observed that Gcn5 HAT activity is required for proper co-transcriptional recruitment of the spliceosome in yeast where U2 snRNP association with the pre-mRNA and subsequent spliceosomal rearrangements are sensitive to Gcn5-dependent histone acetylation. Gcn5-dependent histone acetylation is enriched at the promoter, but the HAT was surprisingly found throughout the intron-containing genes, *DBP2* and *ECM33*. Deletion of histone deacetylases (HDACs) *HOS2* and *HOS3* revealed H3K9 and H3K14 acetylation throughout the gene bodies and abnormal persistence of U2 snRNP factors

near the branch-point region of the pre-mRNA—with aberrant recruitment of downstream splicing factors. While Gcn5-dependent acetylation is important for the fate of spliceosomal rearrangements, the genome-wide implications and overall mechanism underlying this result is not yet clear.

Ribosomal protein genes (RPGs) represent the most abundant subset of intron-containIng genes in any given yeast cell. In a vegetative cell, the pre-mRNA molecules synthesized from RPGs comprise approximately ninety percent of splicing substrates (Ares, 1999; Lopez and Seraphin, 1999; Warner, 1999). In yeast, exit from a proliferative state to entry into meiosis can function as a response to environmental circumstances that call for differentiation of gene expression programs that will reduce ploidy and increase survival (Esposito and Klapholz, 1981; Mitchell, 1988). This transition typically occurs in response to nutrient (such as nitrogen) deplete conditions. The gene expression program governing early meiotic genes responds to and is activated by this nutrient sensitivity and activates a major transcriptional reprogramming event--namely, global repression of RPG transcription (Mitchell et al., 2010; Juneau et al., 2007). Additionally, the efficiency of pre-mRNA specific to the meiotic gene expression program increases the efficiency of splicing during this stage (Juneau et al., 2007; Munding et al., 2013). After the RPGs are repressed in early meiosis, they are subsequently induced during late meiosis, even while starvation conditions persist (Chu et al., 1998; Munding et al., 2010; Primig et al., 2000).

Munding et al discovered that competition between pre-mRNA substrates for limited spliceosomes in the cell contributes to the splicing regulation--particularly in responses to meiosis and cellular stress. Due to the highly abundant and intron-rich nature of the RPGs in yeast, during vegetative growth, spliceosomes are disproportionately titrated away from the remainder of the pre-mRNA pool. Splicing of these transcripts is therefore not as efficient as overall splicing is buffered by the limited availability of spliceosomes. At the onset of meiosis, RPG expression decreases, and global splicing efficiency (including splicing of meiotic transcripts) increases as a result of competition for the limited spliceosomes being relieved.

Here, I show that although Gcn5 is a general transcriptional regulator, it facilitates varied effects on the genome in regards to splicing and expression. These effects have been categorized and with deeper analysis, I found that multiple layers of regulation underlie these multiple outcomes of the loss of *GCN5* in the *S. cerevisiae* genome.

2.2 Materials and methods

Strains and Culture Conditions

All yeast strains used in this study were derived from BY4741 and are listed in the table below. Wildtype, *gcn5Δ* and *H3Δ9-16* yeast strains were grown on solid YPD agar plates and colonies were inoculated into two separate flasks of liquid media per strain (two (2) biological replicates). Each strain was grown to mid-log phase at 30 degrees Celsius in 30mL YPD media

(1% yeast extract, 2% peptone, 2% dextrose) and 10mL was collected via centrifugation for RNA isolation and RNA-seq library prep.

Strain Name	Description	Source
WT	BY4741	This study
gcn5 Δ	*his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0 MATa*	This study
H3 Δ 9-16	MATa *his3 Δ 200 leu2 Δ 0 lys2 Δ 0 trp1 Δ 63 ura3 Δ 0 met15 Δ 0 can1::MFA1pr-HIS3 hht1-hhf1::NatMX4 hht2-hhf2::[HHTS-HHFS]*-URA3 plus pJP11 plasmid [CEN LYS HHF1-HHT1]*	GE Dharmacon

RNA Isolation and RNA-seq Library Preparation

RNA was isolated via hot phenol:chloroform:isoamyl alcohol (PCA) extraction with SDS as described in Ares, 2012 from each biological replicate. Ethanol precipitation at -20 degrees Celsius was performed to precipitate the RNA and concentration was quantified with a NanoDrop 2000 spectrophotometer (ThermoFisher). A total of 20 µg of RNA was treated with DNase I (Roche) and depleted of rRNA with the Ribo-Zero Gold rRNA Removal Kit (Illumina) for each of the 6 RNA samples.

RNA-seq library preparation was performed as described per the Illumina TrueSeq RNA Library Prep Kit v2 protocol. Multiplexed RNA libraries were generated with two (2) biological replicates for each genetic strain (Wildtype, *gcn5 Δ* and *H3 Δ 9-16*)--totaling 6 samples. Final library preparation of each sample was 10nM of 260bp cDNA fragments, as quantified by Qubit. 100 base pair single-end reads were generated on Illumina HiSeq 4000. Sequencing depth was 45948517 and 46236120 for each lane sequenced. Reads were aligned to the SacCer3 genome

assembly using the STAR alignment package (http://code.google.com/p/rna-star/). Genes with less than 5 RPKM in the average of WT conditions were filtered out.

Splicing Efficiency Calculation

Splicing efficiency is calculated via the following ratio equation: [spliced counts]/([spliced counts] + [unspliced raw counts]) where "spliced counts" represents the number of aligned reads flanking exon-exon boundaries of an intron-containing gene and "unspliced counts" represents the number of aligned reads flanking intron-exon, intron-only and exon-intron regions of the same intron-containing gene. The reads in each category are normalized by functional length to account for the number of possible alignments per category. Genes with less than 5 unspliced raw counts in WT were filtered out.

Quantification and Statistical Analysis

Original RNA-seq data represented in all graphs and plots represent the average of 2 biological replicates. For RNA-seq analysis, Spearman's correlation for non-parametric test was used to compute the p-values between the noted x and y data. Error bars in the bar graphs represent the standard error of the mean (SEM). Associated P-values were determined by Mann-Whitney test/one-way ANOVA as stated per figure legend. *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; ****, $p \leq 0.0001$. Statistical analyses were performed in Microsoft Excel (Version 15.28) and Prism 8 (GraphPad).

Data Visualization

Original RNA-seq data was represented in bar graphs while Venn Diagrams and scatter plots were produced in Prism 8 (GraphPad).

2.3 Results

GCN5 and its major histone target residues are required for proper RNA expression of intron-containing transcripts genome-wide

RNA-seq was employed to understand how expression and splicing are affected by Gcn5 or its major histone targets globally. Upon *gcn5Δ* and *H3Δ9-16*, global average RNA expression level was not affected (Fig. 1A). I predicted that these mutants would not affect global average RNA expression as previous literature echoes Gcn5's lack of effect on transcriptional output. To explore whether groups of genes are affected by either mutant in a single direction, the RPKM for all protein coding genes in either wildtype or mutant (filtering parameters: ≥5 RPKM in WT) were plotted on a scatter plot (Fig. 1B and Fig. 1C).

Here, I found that the intron containing class of genes (ICGs) display distinct features in gene expression in response to Gcn5-histone acetyltransferase mutants. To explore whether the RPKM levels of ICGs are behaving the same as all genes, I plotted the average RNA expression level for WT, *gcn5Δ* and *H3Δ9-16* (Fig. 1D and Fig. 1E). Surprisingly, I found that Gcn5-mediated histone acetylation decreases the average expression of ICGs as a functional group (Fig. 1G). To determine if the wildtype expression level of these ICGs is a determinant of how Gcn5-mediated histone acetylation affects RNA expression changes, I grouped all ICGs into quintiles based on WT expression level and plotted these against the change in RPKM level in

the mutant background (Fig. 1F). Gcn5-mediated histone acetylation decreases the average expression of highly expressed ICGs. Additionally, most of the lowly expressed ICGs are benefiting from both *gcn5Δ* and *H3Δ9-16*, while the highly expressed ICGs are decreasing. SInce the spread of these highly expressed ICGs (in regards to change in RPKM level) clump together in a distinct manner, there may be a strong relationship between WT expression level and effect on transcription for these intron containing genes.

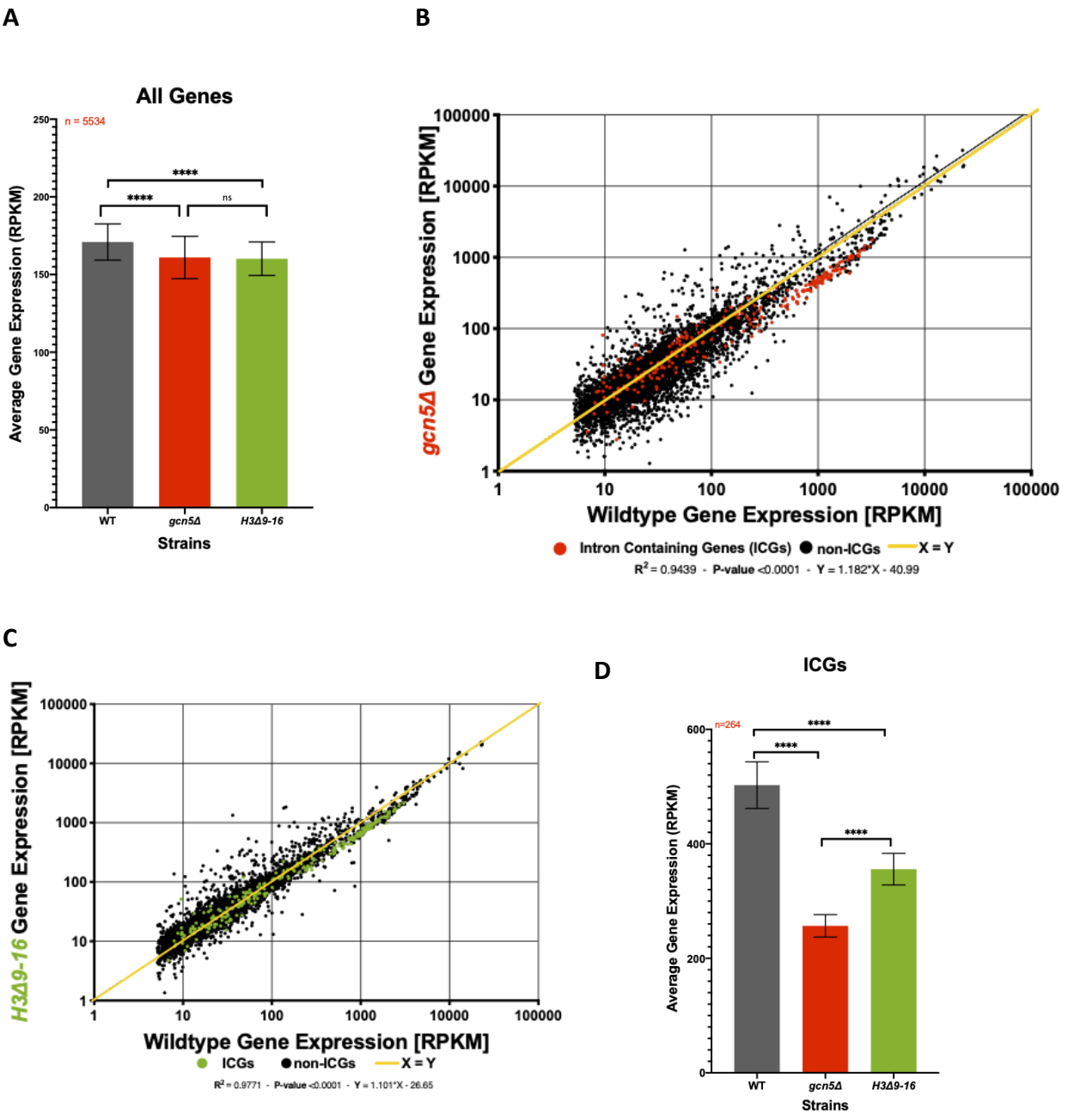


FIGURE 1 (A-D): Gcn5 and its major histone target residues are required for RNA expression of intron containing transcripts - A) Bar graph representing the average RPKM level of all genes (n = 5534 genes) for each strain determined by RPKM values across two biological replicates. SEM is calculated and represented as error bars. B-C) Each scatterplot represents the average mutant RPKM level (y-axis) relative to average WT RPKM level (x-axis). Values represent average across two biological replicates and yellow line represents x=y. Each black dot represents a single non-ICG while red dots in B) are ICGs in *gcn5 Δ* and green dots in C) are ICGs in *H3 Δ 9-16* ($p < 0.0001$). D) Bar graph representing the average RPKM level of ICGs (n = 264) for each strain determined by RPKM values across two biological replicates. SEM is calculated and represented as error bars.

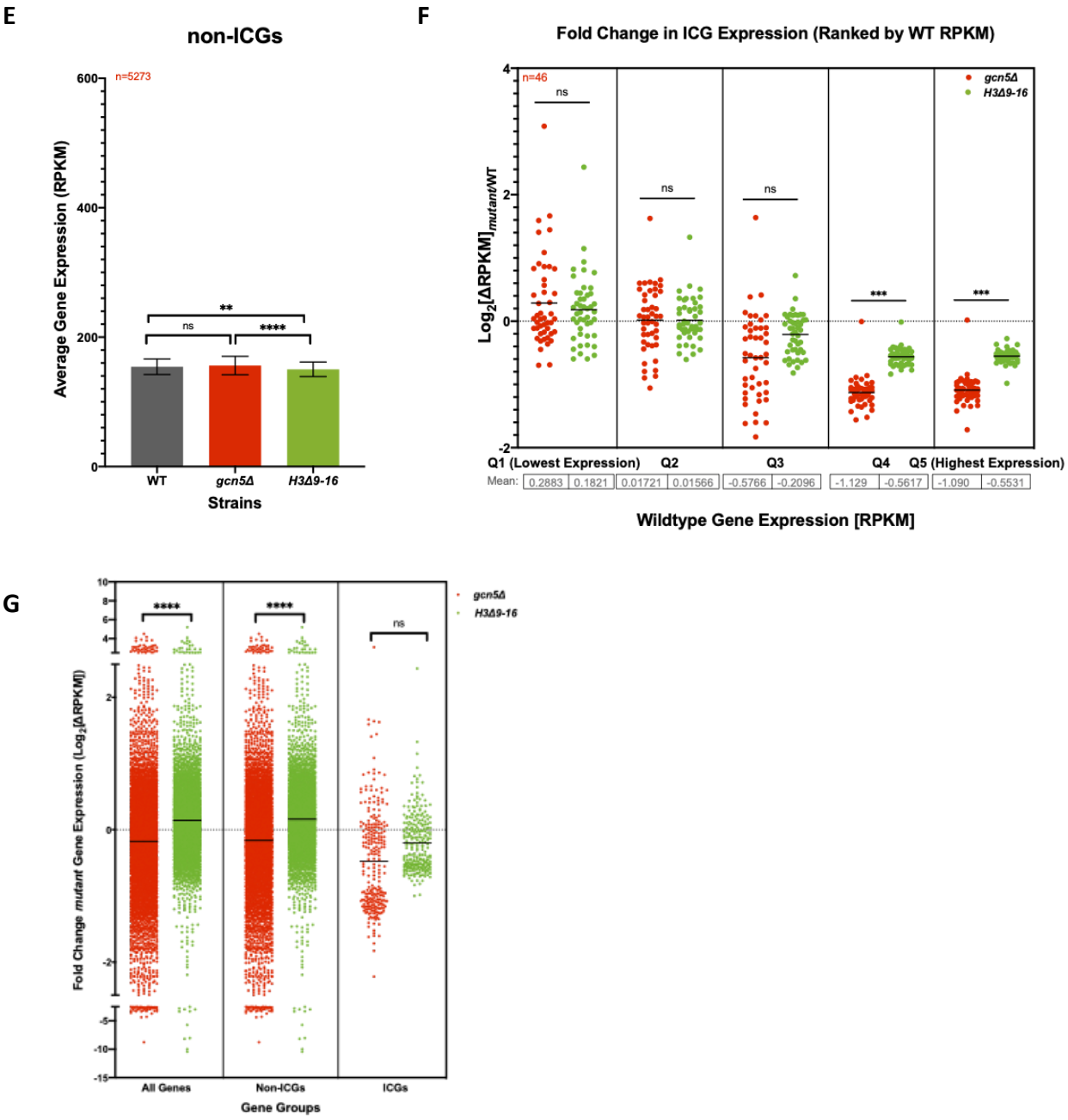


FIGURE 1 (E-G): Gcn5 and its major histone target residues are required for RNA expression of intron containing transcripts - E) Bar graph representing the average RPKM level of all non-ICGs (n = 5273, $p < 0.01$) for each strain determined by RPKM values across two biological replicates. SEM is calculated and represented as error bars. F) Fold change in average mutant RPKM level of ICGs (for both mutants) ranked by average WT RPKM level. ICGs were ranked by WT RPKM level from lowest to highest and categorized into 5 bins of n = 46 genes each. Y-axis represents $\log_2(\Delta \text{RPKM})$ for mutant RPKM relative to WT RPKM for each gene. G) Plot of fold change mutant RPKM level relative to WT for all genes (n = 5534, $p < 0.0001$), non-ICGs (n = 5273, $p < 0.0001$) and ICGs (n = 264, ns) for each mutant.

Expression of intron-containing-RPGs (IC-RPGs) decreases in response to Gcn5 and H3 mutation while expression of IC-non-RPGs are both increasing and decreasing.

RIbosomal protein genes comprise the largest functional class of intron-containing genes in yeast (Fig. 2A). They also are the most highly expressed and well-spliced group of genes. To determine if all ribosomal protein genes are decreasing in expression, I plotted expression changes based on IC-non-RPGs and IC-RPGs categories. Deletion of *GCN5* and the locus harboring the major histone targets of Gcn5p (*H3 Δ9-16*) resulted in the decreased expression of all intron-containing RPGs (IC-RPGs), while non-RPGs are increasing and decreasing in expression (Fig. 2B). Although this result is striking, the extent of relative down-regulation between *gcn5 Δ* and the *H3 Δ9-16* mutant on RPG expression is not surprising, as Gcn5p directly regulates the promoters of RPGs and is still present in the H3 mutant. I also explored the overlap of IC-non-RPGs with similar RNA expression changes between *gcn5 Δ* and *H3 Δ9-16* mutant and found that there is no bias in the overlap towards any direction (increasing or decreasing), and no shared functional group in these genes (Fig. 2C and Fig. 2D).

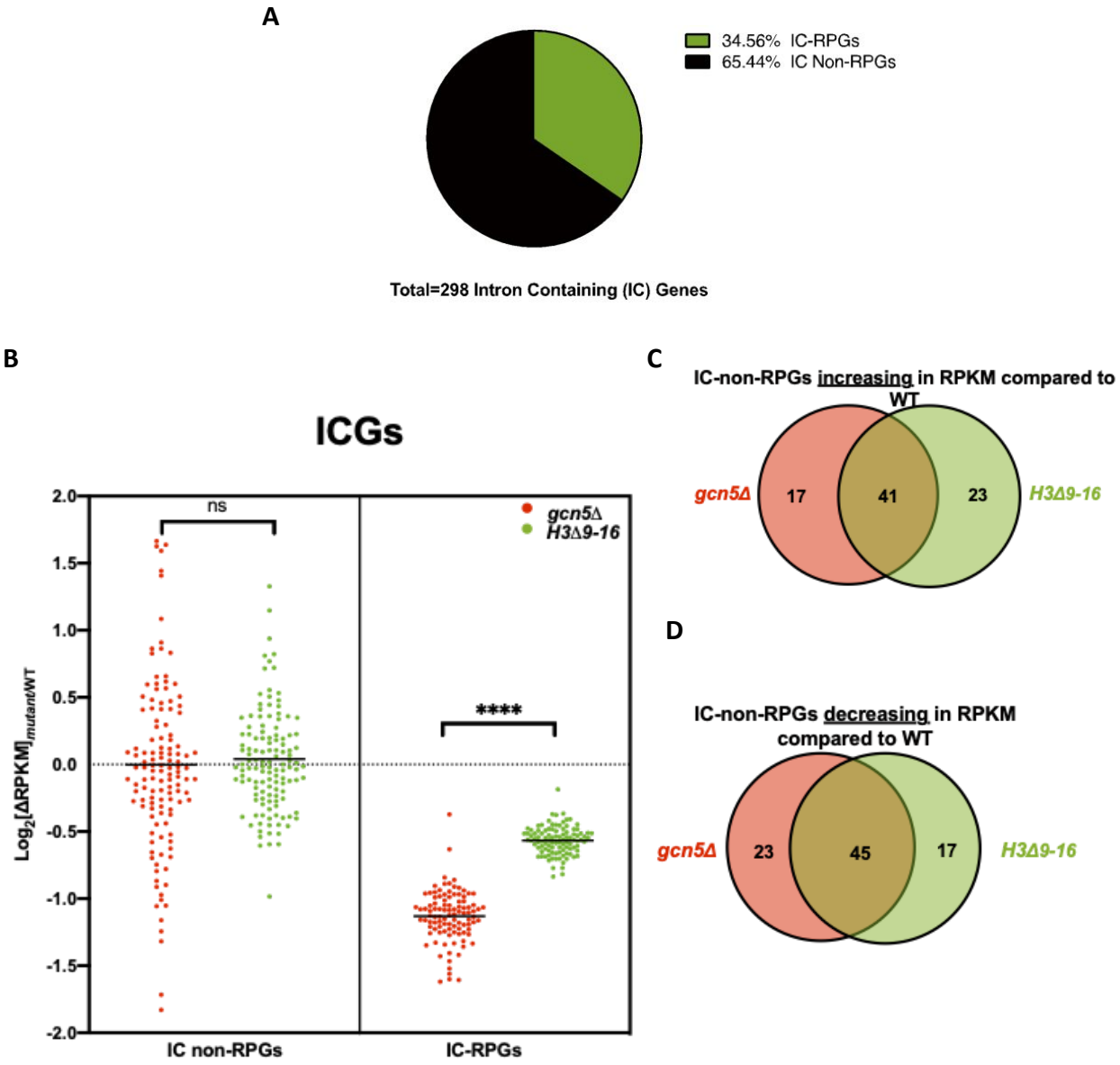


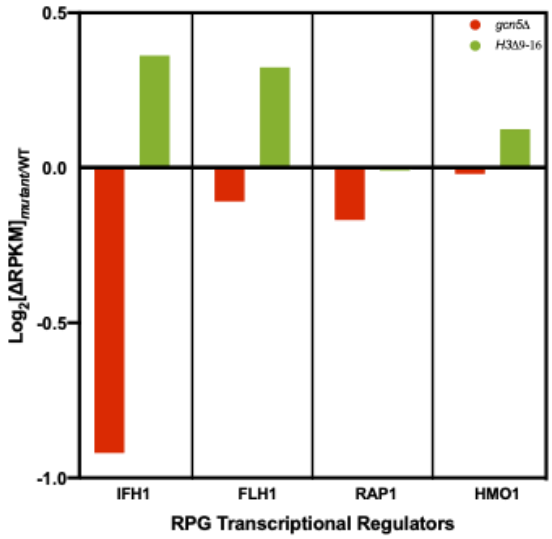
FIGURE 2: Deletion of *GCN5* and its major histone target residues decrease RNA expression of all intron containing ribosomal protein genes - A) Pie-chart representing total number of *S. cerevisiae* intron containing genes with percentage of ribosomal protein genes (RPGs) and non-RPGs. B) Plot representing fold-change in average mutant RPKM level relative to average WT RPKM for IC non-RPGs and IC-RPGs for each mutant strain. C) Venn Diagram of IC-non-RPGs increasing in average mutant RPKM compared to average WT RPKM between mutants (n = 81). D) Venn Diagram of IC-non-RPGs decreasing in average mutant RPKM compared to average WT RPKM between mutants (n = 85).

***RPG expression changes under gcn5 Δ* and *H3 Δ9-16* mutation are not solely due to changes in the expression of RPG transcriptional regulators**

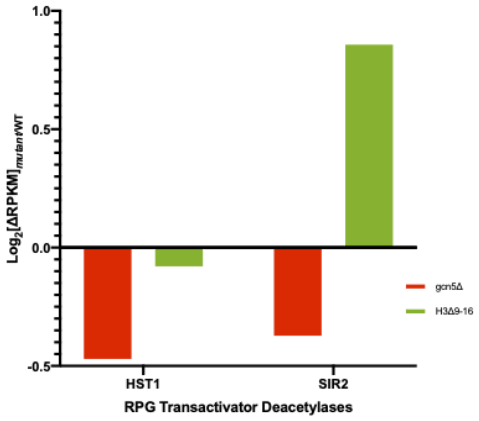
In order to understand how *gcn5 Δ* and *H3 Δ9-16* was functioning in RPG regulation in my data set, it was important to note how the primary factors of RPG regulation were affected by experimental conditions. In RPG regulation, Gcn5 acts to acetylate RP-transactivator, Ifh1 which negatively regulates RPG expression, followed by deacetylation by sirtuins, Sir2 and Hst1 (Downey et al., 2013). If Gcn5-HAT activity was directly responsible for the decrease in RPG expression, the primary factors of RPG expression would be down-regulated under both mutants. *IFH1*, is down-regulated in the *GCN5* mutant and up-regulated in the H3 mutant (Fig. 3A). Other regulators of RP gene expression are unaffected in the mutants.

Both *FHL1* and *RAP1* are decreasing in the *gcn5 Δ* mutant while increasing and decreasing in the *H3 Δ9-16*, respectively (Fig. 3A). Knight et al., 2014 reported that Hmo1 is another regulator of RPG expression, and in the RNA-seq dataset, RPKM levels of *HMO1* are decreasing in the *gcn5 Δ* mutant and increasing in the H3 mutant [data not shown]. Hmo1 is described to toggle the +1 nucleosome from a repressive to active position to regulate RPGs (Reja et al., 2015). Hst1 and Sir2 are the primary deacetylases that deacetylate Ifh1 protein in response to Gcn5 acetylation of RP-promoter-bound Ifh1 and these are both down-regulated in *GCN5* mutants only and slightly up-regulated in H3 (Fig. 3B). Other Gcn5 non-histone targets (*RSC4* and *CSC6*) are decreasing in transcription in *GCN5* mutants but not other mutants (*CSC6* is upregulated in the H3) [data not shown].

A



B



C

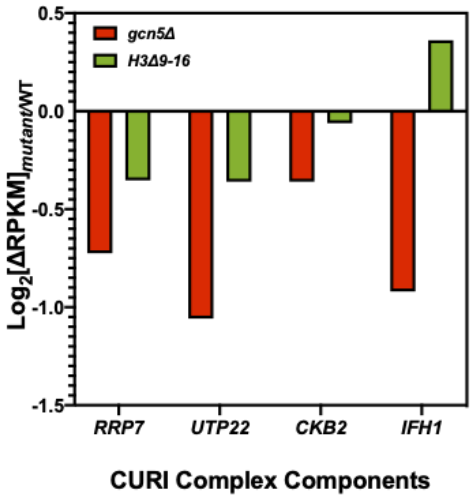


FIGURE 3: RPG expression changes under *gcn5 Δ* and *H3 Δ9-16* mutation are not solely due to changes in the expression of RPG transcriptional regulators - A-C) Fold change in average mutant RPKM level of RPG transcriptional regulators (*IFH1, FLH1, RAP1* and *HMO1*), RPG transactivator deacetylases (*HST1* and *SIR2*) and CURI complex components (*RRP7, UTP22, CKB2* and *IFH1*) in both mutants, respectively. Y-axis represents $\log_2(\Delta \text{RPKM})$ for mutant RPKM relative to WT RPKM for each gene.

Acetylated Ifh1 functions to limit RPG expression in the cell and this modified form of the protein is often found accumulated in the CURI complex (consisting of Rrp7, Utp22 and Ckb2) which functions to link rRNA processing to RPG transcription (Downey et al., 2013; Rudra et al., 2007). In order to rule out the role of *gcn5Δ* and *H3Δ9-16* in modulating RPG expression, I analyzed the expression of CURI complex components (Fig. 3C). Surprisingly, all components of the complex were down-regulated in the absence of *GCN5*, while the H3 mutant down-regulated expression of 3 out of the 4 components. The lack of consensus on RPG factor down-regulation between Gcn5 and lysines 9-16 on H3 suggests that the activity of Gcn5 in histone acetylation is not directly responsible for the overall decrease in RPG expression and other mechanisms may be facilitating this effect.

My data however supports that *GCN5* is required to keep RPGs transcribing at a normal level. The primary histone target region of Gcn5 (Lysines 9 and 14 within the residues 9-16 on the N-terminal tail of Histone H3) is also needed to maintain the normal transcription level of RPGs, even though Gcn5p is still present (which negates the confounding role of Gcn5 in directly acetylating Ifh1--which binds to RPG promoters and negatively regulates their transcription (Downey et al., 2013)). An important consideration to keep in mind is that this result is likely not due to a direct relationship between Gcn5-HAT activity and RPGs, as functional redundancy between TFIID and SAGA co-binding at RPG promoters may be also contributing to this regulation (Huisinga et al., 2004; Lee et al., 2000; Grant and Pugh, 2007).

Ribosomal protein genes are greatly subject to regulation by the TOR kinase pathway, which is one of the major pathways of regulating sensitivity to nutrient availability in yeast. In vegetatitve conditions, the TOR kinase promotes the recruitment of Ifh1, one of the major RPG transactivators, to RPG promoters, where it then interacts with the constitutively promoter-bound RPG transcription factor, Fhl1 and in turn activates expression of RPGs (Martin et al., 2004; Schawalder, 2004; Wade, 2004; Rudra, 2005). Ifh1 recruitment to RPG promoters is essential for RPG expression under tight regulation under nutrient starvation conditions (Martin et al., 2004; Schawalder, 2004; Wade, 2004; Rudra, 2005). Gcn5p is known to acetylate this key RPG transcriptional activator at RPG promoters, which negatively regulates this gene expression program (Downey et al, 2013). Due to the co-regulatory role of Gcn5p in global histone acetylation and acetylation of Ifh1, I am interested in understanding the expression of RPGs in a *GCN5* and *H3* mutant background in order to separate the different modes of regulation by this HAT and further understand how Gcn5-mediated acetylation may affect pre-mRNA splicing.

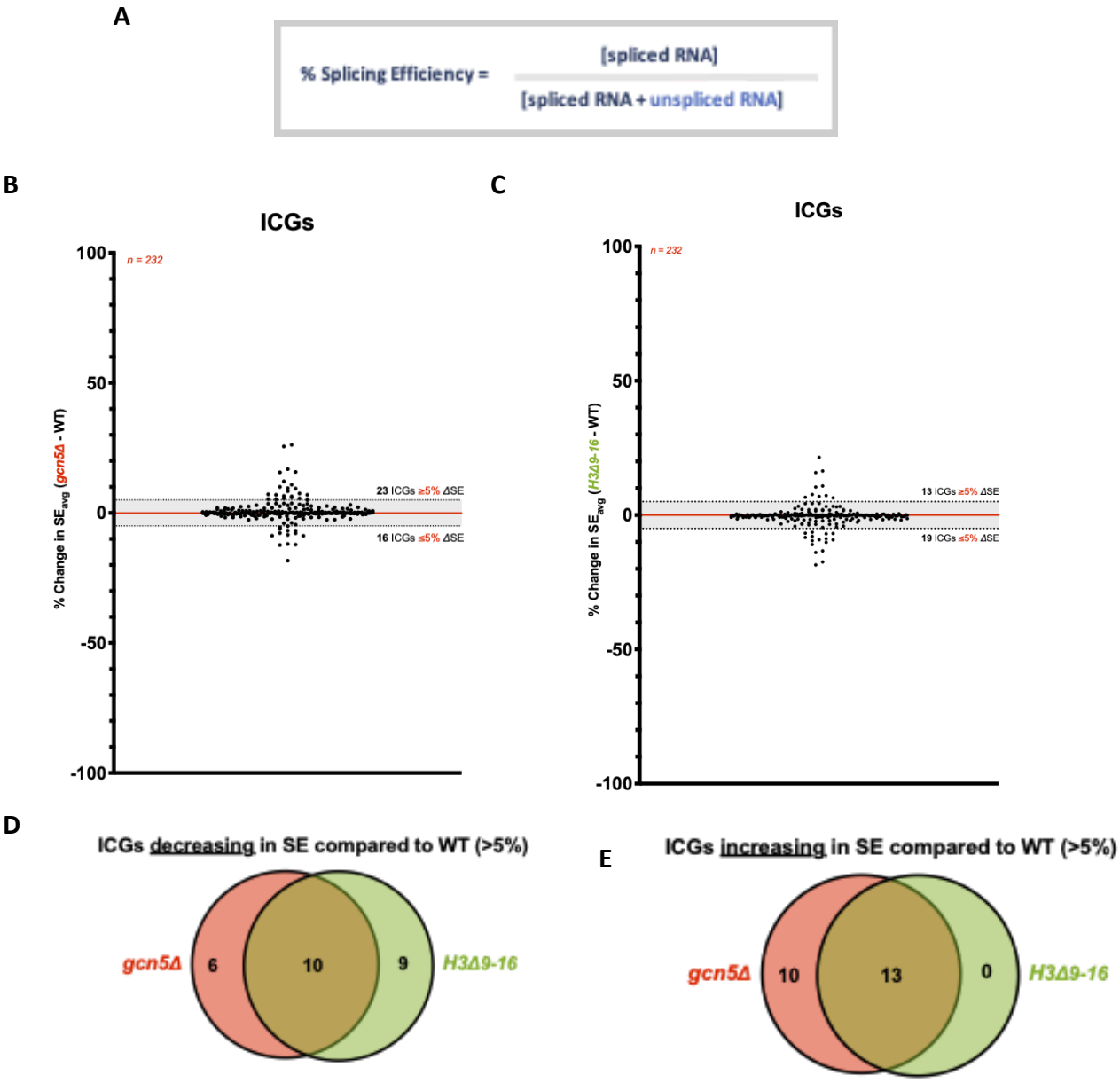
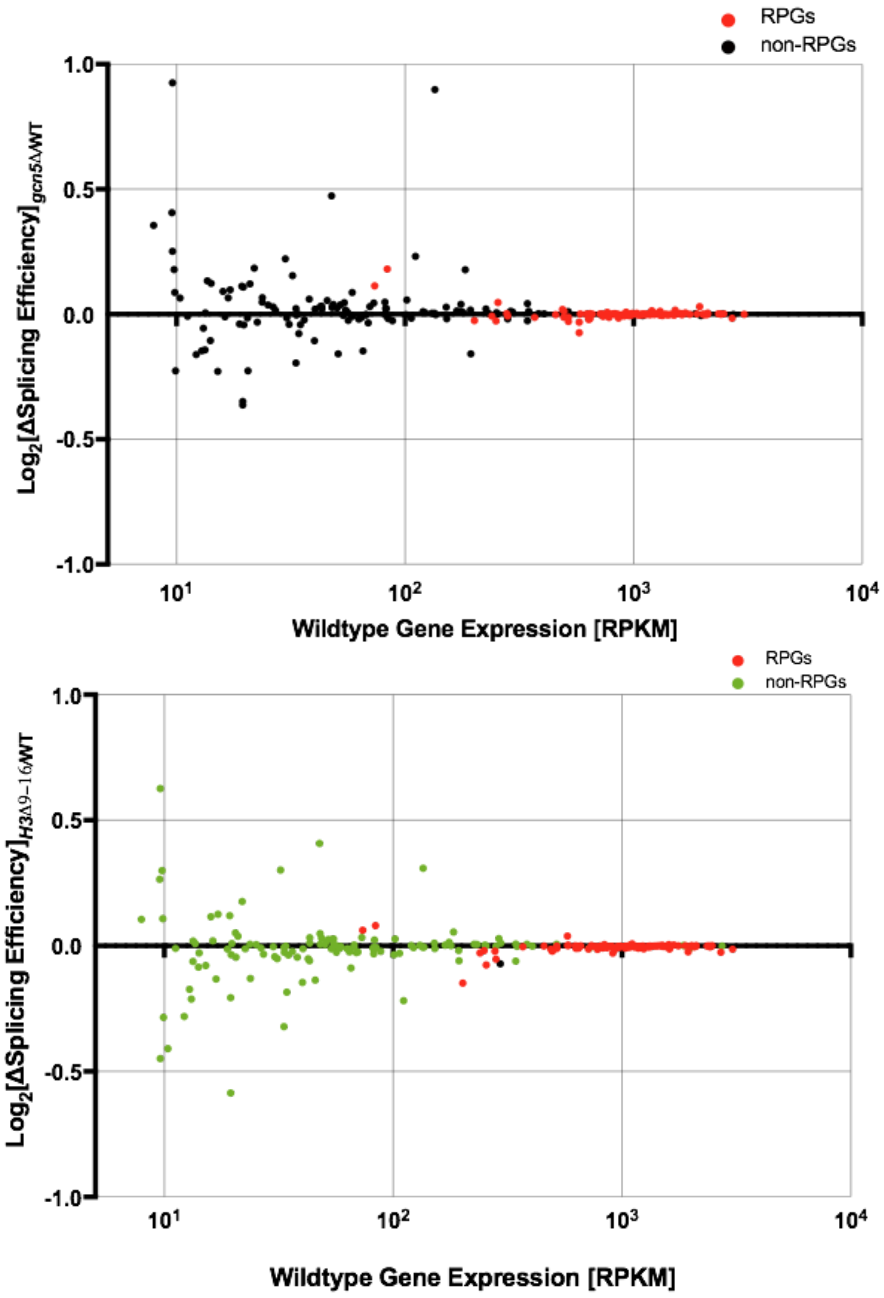


Figure 4 (A-E): *GCN5* and its major histone target residues function in maintaining proper Splicing Efficiency (SE) of ICGs genome-wide - A) Schematic of % splicing efficiency calculation (% Splicing Efficiency = [spliced RNA]/[spliced RNA + unspliced RNA]). B) Plot representing percent difference in Splicing Efficiency (SE) for filtered intron-containing genes under *gcn5 Δ* relative to wildtype. Splicing efficiencies were averaged between two biological replicates for wildtype and mutant and the difference between mutant and wildtype is represented on the graph. Dashed line represents +/- %5 difference ([mutant] - [wildtype]) in SE relative to wildtype (n = 232 genes, spliced/unspliced read count ≥5 in WT and *mutant*). C) Plot representing percent difference in Splicing Efficiency (SE) for filtered intron-containing genes under *H3 Δ 9-16* relative to wildtype. Splicing efficiencies were averaged between two biological replicates for wildtype and mutant and the difference between mutant and wildtype is represented on the graph. Dashed line represents +/- %5 difference ([mutant] - [wildtype]) in SE relative to wildtype (n = 232 genes) D) Venn diagram representing the overlap of ICGs that constitute <5% decrease in SE in *gcn5 Δ* and *H3 Δ 9-16* compared to wildtype. The numbers shown indicate the number of ICGs represented by each mutant after filtering. E) Venn diagram representing the overlap of ICGs that constitute >5% increase in SE in *gcn5 Δ* and *H3 Δ 9-16* compared to wildtype. The numbers shown indicate the number of ICGs



(cont'd) **Figure 4 (cont'd) (F-G):** represented by each mutant after filtering. F) Scatter Plot representing fold change in SE for *gcn5 Δ* relative to wildtype versus wildtype gene expression for all ICGs after filtering (n = 232). Intron-containing Ribosomal Protein Genes (RPGs) are highlighted in red and non-RPGs are highlighted in black. X-axis is represented as wildtype gene expression in RPKM and y-axis is the Log(base2) transformation of the ratio between SE in *gcn5 Δ* and wildtype. G) Scatter Plot representing fold change in SE for *H3 Δ 9-16* relative to wildtype versus wildtype gene expression for all ICGs after filtering (n = 232). Intron-containing Ribosomal Protein Genes (RPGs) are highlighted in red and non-RPGs are highlighted in green. X-axis is represented as wildtype gene expression in RPKM and y-axis is the Log(base2) transformation of the ratio between SE in *H3 Δ 9-16* and wildtype.

Surprisingly, there is a high overlap between genes that are decreasing in splicing efficiency as well as increasing in splicing efficiency between *GCN5* and the H3 mutant, suggesting that the splicing outcomes of these genes is directly influenced by *gcn5 Δ* and *H3 Δ9-16* (Fig. 4D and 4E). Given the overall decreased RPG expression shared in both *gcn5 Δ* and *H3 Δ9-16* as well as the lack of relationship between change in splicing and WT RNA expression outcomes (Fig. 4F and 4G), this high overlap between *gcn5 Δ* and *H3 Δ9-16* amongst IC-non-RPGs improving in SE presents intriguing support for the RPG effect occurring in a *gcn5 Δ* and *H3 Δ9-16* context. This observation is also consistent with other reports regarding the RPG effect (Awad et al., 2017). However, the IC-non-RPGs that are decreasing in SE have yet to be explained by any previously reported mechanism or shared functionality, particularly as there is no enrichment for any particular GO term. The above figures show that there is surprisingly a lack of consensus on the effect of splicing in a single direction for both mutants (Fig. 4A and 4B). Additionally, there lacks a strong relationship between the change in splicing efficiency and level of wild-type gene expression for ICGs as a whole (Fig. 4F and 4G), however, there may be a relationship between functional categories (i.e. RPGs and non-RPGs) and consequences of Gcn5-dependent co-transcriptional splicing. So far, it may be predicted that both mutants would lead to a decrease in co-transcriptional splicing overall given their critical function in transcriptional regulation as well as the results from the previous work in the lab showing aberration of splicing factors on intron containing genes under similar mutations.

Given the surprisingly polarized result, there may be a functional and perhaps regulatory mechanism which is establishing the balance between improved and defective splicing

efficiency. One possibility could be a shift or dysregulation of the total abundance of splicing factors in each mutant cell. Previous work showed that deletion of Gcn5 and perturbation of histone acetyltransferase dynamics resulted in a reduced physical recruitment of two major U2 snRNP components to the branchpoint-adjacent region of two intron-containing genes (Gunderson and Johnson, 2009; Gunderson et al., 2011). I then sought to analyze the expression of splicing factors under each condition.

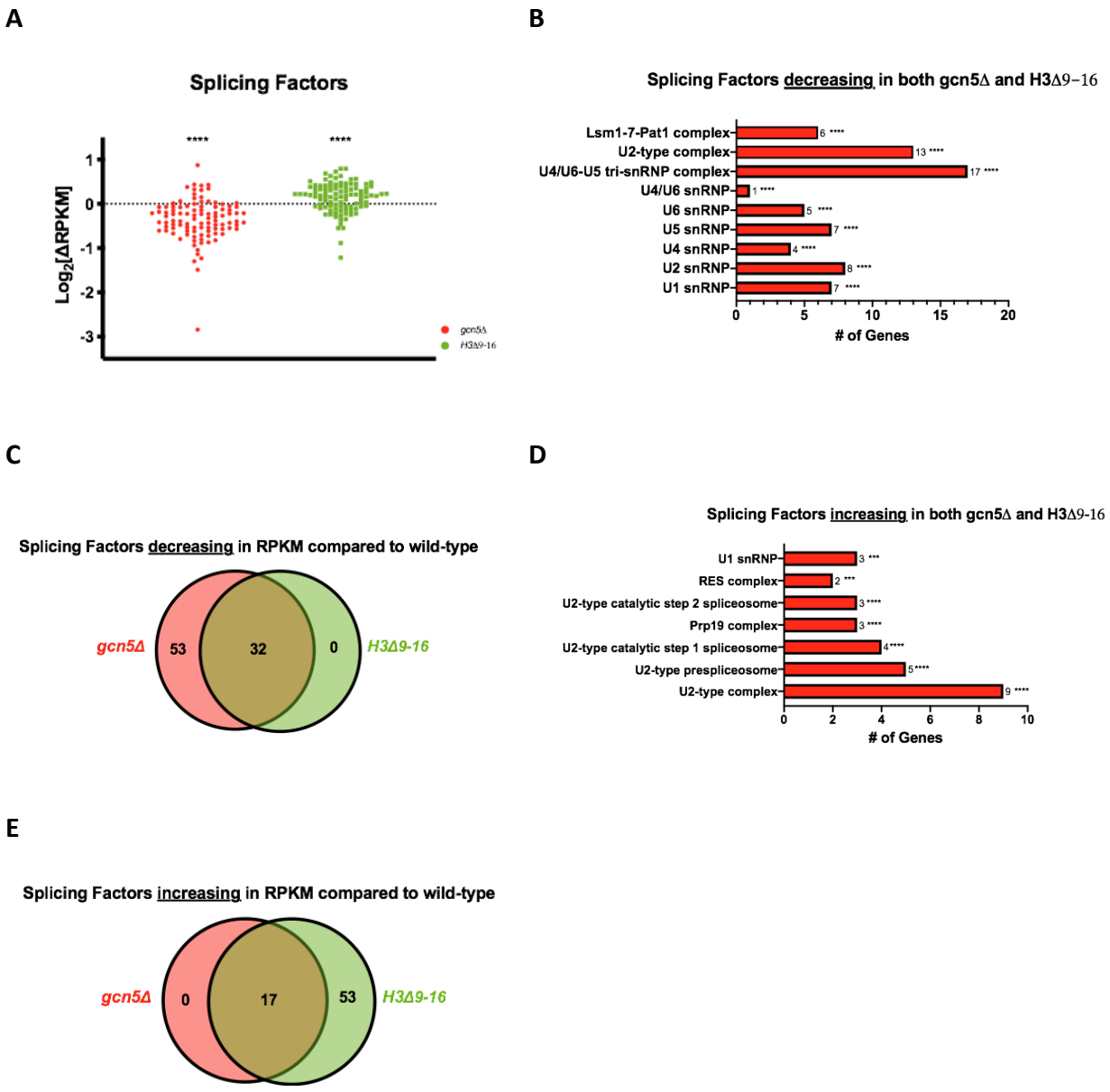


Figure 5: *GCN5* and the H3 lysine 9-16 residues are necessary for maintaining normal RNA expression of splicing factors - A) Plot representing fold change in average RPKM for each mutants relative to wildtype for all splicing factors after filtering (n = 103; RPKM ≥5 counts in WT and *mutant*). List of splicing factor genes was generated from the Saccharomyces Genome Database (SGD) Gene Ontology search term for 'RNA splicing' (GO:0008380). Each gene encoding a splicing factor is represented as *gcn5 Δ* highlighted in red and *H3 Δ 9-16* are highlighted in green. Y-axis is the Log(base2) transformation of the ratio of change in average RPKM level between either *gcn5 Δ* (P-value < 0.0001) or *H3 Δ 9-16* (P-value < 0.0001) and wildtype. B) Gene ontology enrichment of splicing factors decreasing in mutant average RPKM relative to average WT RPKM in both mutants. C) Venn diagram representing the overlap of splicing factor genes that constitute any decrease in RPKM level in *gcn5 Δ* and *H3 Δ 9-16* compared to wildtype. The numbers shown indicate the number of splicing factor genes represented by each mutant. D) Gene ontology enrichment of splicing factors increasing in mutant average RPKM relative to average WT RPKM in both mutants. E) Venn diagram representing the overlap of splicing factor genes that constitute any increase in RPKM level in *gcn5 Δ* and *H3 Δ 9-16* compared to wildtype. The numbers shown indicate the number of splicing factor genes represented by each mutant.

The Venn Diagrams above revealed that 17 splicing factor genes were up-regulated in RPKM expression under both *gcn5 Δ* and *H3 Δ 9-16* compared to 32 genes down-regulated in both mutants (Fig. 5C and 5E). I performed Gene Ontology annotation on this group of overlapped factors to determine if there is a consensus to the step in splicing for which these up- and down-regulation events represent. Interestingly, the highest number of splicing factors down-regulated in both *gcn5 Δ* and *H3 Δ 9-16* are associated with the U4/U6-U5 tri-snRNP complex (17/32 genes)--seconded by the U2 complex (13/32 genes) (Fig. 5B). I found this result intriguing as Gunderson et al, 2011 observed that histone deacetylation is necessary for recruitment of snRNPs downstream of U2. Additionally, as GO analysis presents there to be 34 genes annotated with this complex in SGD, over 50% of the splicing factors down-regulated between *gcn5 Δ* and *H3 Δ 9-16* mutants are represented here. Amongst the U2 complex genes down-regulated between *gcn5 Δ* and *H3 Δ 9-16* mutants was *MSL1*, another gene implicated in the earlier studies with Gcn5p and splicing (Gunderson and Johnson, 2009).

The same analysis as above was performed for the 17 splicing factors sharing up-regulation in expression for both mutants (Fig. 5E) and the component with the largest shared annotation was the U2-type spliceosome complex (9/17 genes) with U2-type pre-spliceosome factors being the more specific annotation (5/17 genes) (Fig. 5D). Combining these two results, an interesting case seems to be further supported for the role of *gcn5 Δ* and *H3 Δ 9-16* in spliceosome assembly as the factors that are upregulated sequentially precede the immediate step in assembly for which the factors that are down-regulated represent. This data suggests that *gcn5 Δ* and *H3 Δ 9-16* (be it direct or indirect) may be primarily functioning during the

transition between U2 snRNP assembly, disassembly and tri-snRNP assembly. The pre-spliceosome complex is implicated in the association of the 5' SS with U1, while the BPS is recognized by the U2 snRNP, prior to the pre-catalytic spliceosome (Will & Luhrmann, 2011).

Expression of intron-containing RPGs (IC-RPGs) decreases in response to Gcn5-HAT mutation and splicing efficiency of non-RPGs are bidirectionally affected

The result of *IFH1* down-regulation being associated with overall decreased RPG gene expression is consistent with data from the Ares group (Munding et al., 2013). This group also presented a new model for understanding the limited nature of the spliceosomes in the cell. During normal conditions, the RPGs make up the largest and most efficiently spliced functional class of introns in the cell--making them the primary target of spliceosomes while non-RPGs are left to compete for the remaining pool of spliceosome (Munding et al., 2013). In conditions where RPG expression is decreased (nutrient starvation, meiosis, etc), these spliceosomes are "freed up" from the missing RPG transcripts and left to benefit the splicing of non-RPG introns--thus increasing splicing efficiency of this group of genes (Munding et al., 2013). Given the overall decrease of RPG transcripts in my dataset, I was interested to observe if this synergistic RPG effect was being reproduced in my dataset. Given that upon RPG down-regulation, the splicing efficiency of IC-non-RPGs should benefit from the newly available pool spliceosomes (Munding et al., 2013), I predicted that all IC-non-RPGs should demonstrate increased splicing efficiencies in both the *gcn5Δ* and *H3Δ9-16.* Surprisingly, the IC-non-RPGs demonstrated both increased and decreased splicing efficiencies in each mutant, suggesting that there is a subset of IC-non-RPGs that are not as sensitive to this RPG effect (Fig. 6A). The

effect of transcription did not have an effect on splicing efficiency for IC-RPGs and IC-non-RPGs under either mutant (Fig. 6B-6E). The non-parametric Mann-Whitney test was used to describe the significant difference between the median of IC-RPGs and IC-non-RPGs for each mutant without assuming Gaussian/normal distribution while 2-way ANOVA was used to compare the difference in the combined means of IC-RPGs and IC-non-RPGs between each mutant. Wilcoxon-signed-rank test was used to determine the significance of whether the median SE is greater or lesser than 0 for each subcolumn (IC-RPGs and IC-non-RPGs).

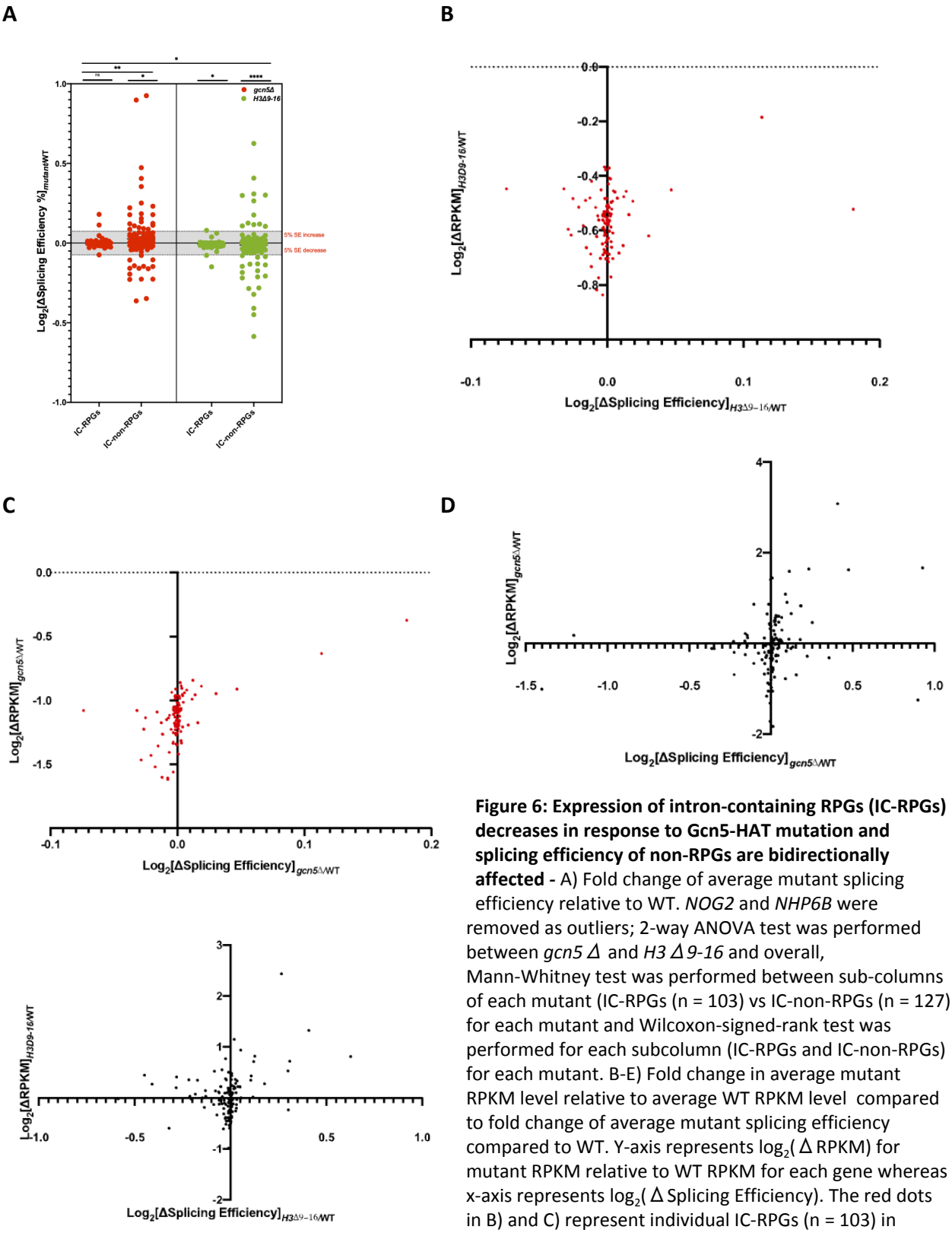


Figure 6: Expression of intron-containing RPGs (IC-RPGs) decreases in response to Gcn5-HAT mutation and splicing efficiency of non-RPGs are bidirectionally affected - A) Fold change of average mutant splicing efficiency relative to WT. *NOG2* and *NHP6B* were removed as outliers; 2-way ANOVA test was performed between *gcn5 Δ* and *H3 Δ 9-16* and overall, Mann-Whitney test was performed between sub-columns of each mutant (IC-RPGs (n = 103) vs IC-non-RPGs (n = 127) for each mutant and Wilcoxon-signed-rank test was performed for each subcolumn (IC-RPGs and IC-non-RPGs) for each mutant. B-E) Fold change in average mutant RPKM level relative to average WT RPKM level compared to fold change of average mutant splicing efficiency compared to WT. Y-axis represents $\log_2(\Delta \text{RPKM})$ for mutant RPKM relative to WT RPKM for each gene whereas x-axis represents $\log_2(\Delta \text{Splicing Efficiency})$. The red dots in B) and C) represent individual IC-RPGs (n = 103) in *H3 Δ 9-16* and *gcn5 Δ*, respectively whereas in black dots in D) and E) represent individual IC-non-RPGs (n = 127) in *gcn5 Δ* and *H3 Δ 9-16*, respectively.

2.4 Discussion

Upon beginning this project, the Johnson lab established that acetylation by the transcriptional coactivator Gcn5 plays a novel role in co-transcriptional spliceosome assembly (Gunderson et al., 2009) and that dynamic histone acetylation is critical for cotranscriptional spliceosome assembly and spliceosomal rearrangements (Gunderson and Johnson, 2011). This data unearthed new ground for understanding the role of chromatin in splicing and fostered new frontiers in this corner of the field. However, the primary gap in knowledge was with regards to the role of Gcn5 in splicing genome-wide. Here, I demonstrated that Gcn5-HAT activity down-regulates RPG expression which increases the availability of limited spliceosomes to improve splicing of a subset of non-RPGs genome-wide.

The results with the histone H3 mutant (*H3 Δ 9-16*) is most intriguing because according to the literature, Gcn5 has a non-histone target that specifically regulates gene regulation of RPGs. In the histone mutant, Gcn5 is free to function in this context (of acetylating Ifh1, which places a "control brake" on aberrant transcription of RPGs in normal contexts, and helps to control/titrate transcription in a nutrient-depleted/stress context) which should predictably not hamper the expression of RPGs. The observation that RPGs are all decreasing in transcription in the histone mutant when Gcn5 is still present to regulate the non-histone target at RPG promoters suggests that acetylation of Histone H3 by Gcn5 is required for RPG expression. There seems to be a role for Gcn5-mediated histone acetylation in positively regulating gene expression of RPGs specifically, as non-RPGs are both increasing and decreasing in expression in

the H3 and Gcn5 mutant. When Gcn5 is removed, the RPKM of all IC-RPGs (with the exception of 1-2) are decreasing.

This work has established a new conceptual framework to consider while exploring questions regarding co-transcriptional splicing mechanisms in yeast as well as general, non-yeast splicing. This data begins to address a systems-level view of how splicing regulatory networks can be interconnected with other gene regulatory processes. Additionally, it proposes an intriguing mechanism by which the cell responds to stress. Coordinating RPG expression with splicing outcomes may be a way of tuning the cells energy in response to the shift in the gene expression needs of the cell. This finding is of particular importance to this dissertation thesis work as Gcn5 is implicated in RPG regulation at the level of chromatin structure and transcription (Downey et al., 2013). Taking into account the competitive strategy described above is crucial to gaining proper conceptual understanding of the biological consequences of Gcn5 in splicing. I aimed to use the findings above as fundamental considerations to guide the experiments in order to understand the effect of Gcn5-dependent acetylation on splicing outcomes--separate from its co-regulatory role in RPG expression--and through the context of stress regulation.

The loss of Gcn5 affects Ifh1 regulation of Ribosomal Protein Genes and seems to compound the with the effect of Gcn5 in gene activation via histone acetylation, leading to a ~50% loss of RPKMs in this functional class relative to wildtype. The loss of Lysines 9 through 16 on Histone H3 may contribute to the decreased ability for the RPGs to activate. In addition to

Gcn5-non-histone effects, these RPG may be highly sensitive to *gcn5Δ* and *H3Δ9-16* as their promoter architecture commands labile nucleosomes (Knight et al., 2014) and available HAT activity to keep transcription at normal levels.

REFERENCES

Ares, M., Jr., Grate, L., & Pauling, M. H. (1999). A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA, 5*(9), 1138-1139. doi:10.1017/s1355838299991379

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O., & Herskowitz, I. (1998). The transcriptional program of sporulation in budding yeast. *Science, 282*(5389), 699-705. doi:10.1126/science.282.5389.699

Downey, M., Knight, B., Vashisht, A. A., Seller, C. A., Wohlschlegel, J. A., Shore, D., & Toczyski, D. P. (2013). Gcn5 and sirtuins regulate acetylation of the ribosomal protein transcription factor Ifh1. *Curr Biol, 23*(17), 1638-1648. doi:10.1016/j.cub.2013.06.050

Huisinga, K. L., & 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*(4), 573-585. doi:10.1016/s1097-2765(04)00087-5

Huisinga, K. L., & Pugh, B. F. (2007). A TATA binding protein regulatory network that governs transcription complex assembly. *Genome Biol, 8*(4), R46. doi:10.1186/gb-2007-8-4-r46

Juneau, K., Palm, C., Miranda, M., & Davis, R. W. (2007). High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. *Proc Natl Acad Sci U S A, 104*(5), 1522-1527. doi:10.1073/pnas.0610354104

Knight, B., Kubik, S., Ghosh, B., Bruzzone, M. J., Geertz, M., Martin, V., . . . Shore, D. (2014). Two distinct promoter architectures centered on dynamic nucleosomes control ribosomal protein gene transcription. *Genes Dev, 28*(15), 1695-1709. doi:10.1101/gad.244434.114

Lee, T. I., Causton, H. C., Holstege, F. C., Shen, W. C., Hannett, N., Jennings, E. G., . . . Young, R. A. (2000). Redundant roles for the TFIID and SAGA complexes in global transcription. *Nature, 405*(6787), 701-704. doi:10.1038/35015104

Lopez, P. J., & Seraphin, B. (1999). Genomic-scale quantitative analysis of yeast pre-mRNA splicing: implications for splice-site recognition. *RNA, 5*(9), 1135-1137. doi:10.1017/s135583829999091x

Malone, R. E., & Esposito, R. E. (1981). Recombinationless meiosis in Saccharomyces cerevisiae. *Molecular and Cellular Biology, 1*(10), 891-901. doi:10.1128/mcb.1.10.891

Martin, D. E., Soulard, A., & Hall, M. N. (2004). TOR regulates ribosomal protein gene expression via PKA and the Forkhead transcription factor FHL1. *Cell, 119*(7), 969-979. doi:10.1016/j.cell.2004.11.047

Mitchell, A. P. (1988). Two switches govern entry into meiosis in yeast. *Prog Clin Biol Res, 267*, 47-66.

Mitchell, S. F., Walker, S. E., Algire, M. A., Park, E. H., Hinnebusch, A. G., & Lorsch, J. R. (2010). The 5'-7-methylguanosine cap on eukaryotic mRNAs serves both to stimulate canonical translation initiation and to block an alternative pathway. *Mol Cell, 39*(6), 950-962. doi:10.1016/j.molcel.2010.08.021

Munding, E. M., Igel, A. H., Shiue, L., Dorighi, K. M., Trevino, L. R., & Ares, M., Jr. (2010). Integration of a splicing regulatory network within the meiotic gene expression program of Saccharomyces cerevisiae. *Genes Dev, 24*(23), 2693-2704. doi:10.1101/gad.1977410

Munding, E. M., Shiue, L., Katzman, S., Donohue, J. P., & Ares, M., Jr. (2013). Competition between pre-mRNAs for the splicing machinery drives global regulation of splicing. *Mol Cell, 51*(3), 338-348. doi:10.1016/j.molcel.2013.06.012

Primig, M., Williams, R. M., Winzeler, E. A., Tevzadze, G. G., Conway, A. R., Hwang, S. Y., . . . Esposito, R. E. (2000). The core meiotic transcriptome in budding yeasts. *Nat Genet, 26*(4), 415-423. doi:10.1038/82539

Reja, R., Vinayachandran, V., Ghosh, S., & Pugh, B. F. (2015). Molecular mechanisms of ribosomal protein gene coregulation. *Genes Dev, 29*(18), 1942-1954. doi:10.1101/gad.268896.115

Rudra, D., Mallick, J., Zhao, Y., & Warner, J. R. (2007). Potential interface between ribosomal protein production and pre-rRNA processing. *Mol Cell Biol, 27*(13), 4815-4824. doi:10.1128/MCB.02062-06

Rudra, D., Zhao, Y., & Warner, J. R. (2005). Central role of Ifh1p-Fhl1p interaction in the synthesis of yeast ribosomal proteins. *EMBO J, 24*(3), 533-542. doi:10.1038/sj.emboj.7600553

Schawalder, S. B., Kabani, M., Howald, I., Choudhury, U., Werner, M., & Shore, D. (2004). Growth-regulated recruitment of the essential yeast ribosomal protein gene activator Ifh1. *Nature, 432*(7020), 1058-1061. doi:10.1038/nature03200

Wade, J. T., Hall, D. B., & Struhl, K. (2004). The transcription factor Ifh1 is a key regulator of yeast ribosomal protein genes. *Nature, 432*(7020), 1054-1058. doi:10.1038/nature03175

Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci, 24*(11), 437-440. doi:10.1016/s0968-0004(99)01460-7

Will, C. L., & Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb Perspect Biol, 3*(7). doi:10.1101/cshperspect.a003707

CHAPTER 3 - The role of Gcn5 in connecting chromatin architecture to splicing of non-RPGs, genome-wide

3.1 Introduction

Nucleosomes are 147bp of DNA wrapped around a histone octamer with 2 copies of H2A, H2B, H3 and H4. Packaging of genomic DNA into chromatin regulates access of proteins that activate gene expression, DNA replication, recombination and repair. Biochemical modification of the histones regulate the state of the chromatin which dictates transcriptional activity. For example, hyper-acetylated chromatin facilitates a transcriptionally active region while hypo-acetylated chromatin does the opposite (a repressed chromatin state). Acetylation of histone H3 and H4 tends to peak at promoters and as thus, Gcn5 and Esa1, two histone acetyltransferases are often found at transcriptionally active promoters (Pokholok et al., 2005; Robert et al., 2004).

Chromatin is the principal template for transcription regulation and it is well-established that the organization of these DNA-protein complexes have vast implications for the physical and kinetic regulation of genes and transcriptional output across the genome. For example, intricate higher order folding of chromatin into Topologically-Associated Domains (TADs) and other looping structures facilitate physical long-range genome interactions between regulatory elements in mammalian cells to enhance or suppress gene expression (Dixon et al., 2012). Post-translational modifications (such as monoubiquitination) of histone proteins within chromatin influence the rate of transcription *in vitro* (Pavri et al., 2005). Given the spatially complex yet systematic nature of transcriptional activation, elongation and termination, abundant opportunities for coordinated regulation can occur between chromatin, transcription and other gene regulatory compartments, such as pre-mRNA splicing.

Pre-mRNA splicing of intron-containing genes (ICGs) occurs by stepwise assembly of the large multi-subunit ribonucleoprotein spliceosome complex onto specific splice site sequences in the intron. This stepwise process of spliceosome assembly onto pre-mRNA occurs co-transcriptionally. While the nascent RNA protrudes from the RNA Pol II exit tunnel during transcription, the spliceosome begins assembling on the transcript almost immediately as the intron is transcribed (Oesterreich et al., 2016). Approximately 40nt from the 5' end of the nascent RNA is the minimum distance required to initiate assembly of the catalytic spliceosome during Step I recognition of the branchpoint sequence in budding yeast (Liu and Cheng, 2012). Considering that the average rate of transcription elongation of budding yeast protein-coding genes is approximately 1.5kb/minute with the full splicing reaction being completed in just under 5 seconds, deciphering the functional and spatio-temporal relationship between spliceosome assembly and the transcriptional environment is of active investigation in the field (Mason & Struhl, 2005; Oesterreich et al., 2016). Given the added dynamicity of the chromatin environment regulating transcription, the relationship between splicing and chromatin is of particular interest.

Recently, there has been a growing appreciation of the various models for the relationship between chromatin and splicing. Direct models tend to describe physical recruitment mechanisms between chromatin factors and spliceosome assembly as well as the kinetic relationship between transcription elongation rate and spliceosome assembly. On the other hand, indirect models, such as the shifting competition of spliceosomes in response to changes

in ribosomal protein gene expression, can affect global splicing outcomes. It is well understood that Gcn5 impacts nucleosomal dynamics and chromatin accessibility and that this can influence transcription. The Johnson lab has shown that chromatin modifiers make unique contributions to pre-mRNA splicing through these direct and indirect models (Leung et al., 2019; Neves et al., 2017; Awad et al., 2017). The data presented in Chapter 2 is the first genome-wide investigation of Gcn5-HAT activity in pre-mRNA splicing and these results suggest an indirect model of regulating splicing outcomes. In particular, the observation of decreased RPG expression followed by increased splicing efficiency of a subset of non-RPGs brings the indirect model of ribosomal protein gene (RPG) regulation into a global splicing context.

One of the ways that splicing of ICGs can be broadly affected is through the availability of spliceosomes in the cell. Ares and colleagues mechanistically demonstrated that spliceosomes represent a limited cellular resource and by virtue, are subject to competition by an array of ICGs (Munding et al., 2013). Given that RPGs represent the largest, most transcript-abundant and most efficiently spliced class of ICGs in the cell, these tend to have a large competitive advantage for sequestering most of the limited spliceosomes in the cell. Non-RPGs are thus left with the remainder of this pool of spliceosomes which sensitizes their splicing outcomes to the fate of RPG expression. In the event that RPG expression decreases, the competition for the limited spliceosomes is relinquished, and non-RPG splicing efficiency benefits. Decreased RPG expression leads to an increase in splicing efficiency due to the increased availability of spliceosomes. The RNA-seq data presented in Chapter 2 demonstrates the consequential relationship between *gcn5Δ* and *H3Δ9-16*, RPG regulation and pre-mRNA splicing in a

genome-wide context. Following the well-established model reported above, in a Gcn5-HAT deficient background, all intron-containing non-RPGs are predicted to benefit from the increased availability of spliceosomes with an increased splicing efficiency.

Surprisingly, a subset of non-RPGs from the RNA-seq data within this model show a decreased splicing efficiency despite the increased availability of spliceosomes. This group of genes may be governed by a separate layer of gene regulation that impacts splicing outcomes. Given the role of Gcn5 in post-translational modification of the chromatin landscape, I hypothesized that there may be a link between chromatin architecture and splicing outcomes at these genes. Potentially, distinct variability in the chromatin accessibility landscape of this group of genes may mediate a lack of sensitivity to the RPG effect. Given the comprehensive evidence implicating the role of chromatin architecture in spliceosome assembly and splicing outcomes (Leung et al., 2019; Matveeva et al., 2019; Henriques et al., 2013; Jimeno-Gonzalez et al., 2015; Andersson et al., 2009; Tilgner et al., 2009; Schwartz et al., 2009; Kolasinska-Zwierz et al., 2009; Jonkers et al., 2014), this hypothesized chromatin landscape variability may facilitate kinetic or physical regulatory processes which desensitize the genes to regulation by the RPG effect. The purpose of this chapter is to identify and characterize this group of genes in order to understand the functional significance for why they lack sensitivity to this RPG effect.

Is there a linear relationship between how much a gene is expressed and if this affects the chromatin landscape in the presence or absence of Gcn5? I predicted that the more expressed a gene is, the less nucleosomes that stably occupy that particular gene locus. There may be a

strong relationship between H3 occupancy and expression. The primary question here is whether there is a connection between this observation and splicing outcomes. I initially sought to understand if splicing outcomes of a gene can be predicted from the landscape of the chromatin environment of the gene. This understanding can be sought from analyzing meta-gene sequencing profiles which describe the enrichment of a few key chromatin marks describing the nature of the loci.

My first approach is to explore the chromatin state of different categories of genes--each category defined by the unique effect of *gcn5 Δ* and *H3 Δ 9-16* on splicing of introns in those categories from Chapter 2. If there is a relationship between the chromatin state of a gene and the effect that *gcn5 Δ* and *H3 Δ 9-16* has on its splicing, then this supports a direct role for Gcn5-mediated HAT activity in genome-wide splicing.

3.2 Materials and Methods

Strains and Culture Conditions

All yeast strains used in this study were derived from BY4741 and are listed in the table below. Wildtype, *gcn5 Δ* and *H3 Δ 9-16* yeast strains were grown on solid YPD agar plates and colonies were inoculated into two separate flasks of liquid media per strain (two (2) biological replicates). Each strain was grown to mid-log phase at 30 degrees Celsius in 30mL YPD media (1% yeast extract, 2% peptone, 2% dextrose) and 10mL was collected via centrifugation for RNA isolation and RNA-seq library prep.

Strain Name	Description	Source
WT	BY4741	This study
gcn5 Δ	*his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0 MATa*	This study
H3 Δ 9-16	MATa *his3 Δ 200 leu2 Δ 0 lys2 Δ 0 trp1 Δ 63 ura3 Δ 0 met15 Δ 0 can1::MFA1pr-HIS3 hht1-hhf1::NatMX4 hht2-hhf2::[HHTS-HHFS]*-URA3 plus pJP11 plasmid [CEN LYS HHF1-HHT1]*	GE Dharmacon

RNA Isolation and RNA-seq Library Preparation

RNA was isolated via hot phenol:chloroform:isoamyl alcohol (PCA) extraction with SDS as described in Ares, 2012 from each biological replicate. Ethanol precipitation at -20 degrees Celsius was performed to precipitate the RNA and concentration was quantified with a NanoDrop 2000 spectrophotometer (ThermoFisher). A total of 20 µg of RNA was treated with DNase I (Roche) and depleted of rRNA with the Ribo-Zero Gold rRNA Removal Kit (Illumina) for each of the 6 RNA samples.

RNA-seq library preparation was performed as described per the Illumina TrueSeq RNA Library Prep Kit v2 protocol. Multiplexed RNA libraries were generated with two (2) biological replicates for each genetic strain (Wildtype, *gcn5 Δ* and *H3 Δ 9-16*)--totaling 6 samples. Final library preparation of each sample was 10nM of 260bp cDNA fragments, as quantified by Qubit. 100 base pair single-end reads were generated on Illumina HiSeq 4000. Sequencing depth was 45948517 and 46236120 for each lane sequenced. Reads were aligned to the SacCer3 genome assembly using the STAR alignment package (http://code.google.com/p/rna-star/).

Splicing Efficiency Calculation

Splicing efficiency is calculated via the following ratio equation: [spliced counts]/([spliced counts] + [unspliced raw counts]) where “spliced counts” represents the number of aligned reads flanking exon-exon boundaries of an intron-containing gene and “unspliced counts” represents the number of aligned reads flanking intron-exon, intron-only and exon-intron regions of the same intron-containing gene. The reads in each category are normalized by functional length to account for the number of possible alignments per category.

Quantification and Statistical Analysis

Original RNA-seq data represented in all graphs and plots represent the average of 2 biological replicates. For RNA-seq analysis, Spearman’s correlation for non-parametric test was used to compute the p-values between the noted x and y data. Error bars in the bar graphs represent the standard error of the mean (SEM). Associated P-values were determined by Mann-Whitney test/one-way ANOVA as stated per figure legend. *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; ****, $p \leq 0.0001$. Statistical analyses were performed in Microsoft Excel (Version 15.28) and Prism 8 (GraphPad).

Public Datasets

The datasets used for meta-gene analysis of Gcn5-AA ChIP-seq (RNAPII-Ser5p, histone H3, H3K9ac) and MNase-seq were derived from Bruzzone et al., 2018. The data files were downloaded publically from the Gene Expression Omnibus under accession series GSE109235

and represent the average of three biological replicates each. Antibodies used for experimental design are as follows: RNAPII-Ser5p (Abcam ab5131), histone H3 (Abcam ab1791) and H3K9ac (Millipore 07-392)

Data Visualization

Original RNA-seq data represented in bar graphs, Venn Diagrams and scatter plots were produced in Prism 8 (GraphPad). For meta-gene analysis, all plots were visualized in SeqPlots interactive software (Version 3.10; Stempor and Ahringer, 2014).

Meta-gene Analysis

All meta-gene analysis was performed by SeqPlots interactive software to determine the average pattern distribution and density of chromatin signatures across genomic loci. The custom BED files used to categorize non-RPGs were generated via the UCSC Table Browser (Karolchik et al., 2004). For 'SE up', all filtered non-RPGs (total = 129 genes) with an average increase in splicing efficiency under both *gcn5Δ* and *H3Δ9-16* (total = 45 genes) were compiled in one custom BED file with all associated genomic regions of interest. The same process was performed for 'SE down', where a total of 44 genes were compiled into a seperate BED file. WIG files with the average occupancy ratio of associated ChIP-seq and MNase-seq data were downloaded from publically available genomic datasets (as described above). Plots analyzed average occupancy ratio for all marks annotated in the corresponding figures and centered on the transcription start site (TSS), with the range of -500bp to +1200bp.

3.3 Results

Wildtype RNA expression level and splicing efficiency does not determine Gcn5-HAT dependent splicing outcomes of IC-non-RPGs

Gcn5-mediated acetylation is increased on highly expressed genes to facilitate chromatin accessibility for transcription (Xue-Franzen et al., 2015). Given the co-transcriptional nature of splicing, I aimed to identify the WT transcription level of IC-non-RPGs that are either increasing or decreasing in both *GCN5* and H3 mutants. If the splicing outcomes of IC-non-RPGs show a distinguishable association to wildtype RNA expression level, then this may suggest the role of RNA expression in sensitizing genes to the RPG effect. Although the data in Chapter 2 demonstrated that there is no relationship between the Gcn5-dependent change in RNA expression level and change in splicing efficiencies for these IC-non-ICGs, the WT level of RNA expression of these IC-non-ICGs may be associated with the splicing outcome under *gcn5Δ* and *H3Δ9-16* mutation. There is a lacking relationship between WT expression level and Gcn5-dependent splicing outcomes (Fig. 1A-D). My next question was to determine if there is a relationship between WT splicing efficiency of these non-RPGs genes and change in splicing efficiency. Are more poorly spliced non-RPGs likely to be decreasing in splicing efficiency? Upon repeating the same analysis for wildtype splicing efficiency and Gcn5-dependent splicing outcomes, there was also a lack of correlation Fig. 1E. The same analysis was repeated for genes with greater than 5% change in splicing efficiency to rule out potential noise of minimal effects on splicing Fig. 1F**.** To rule out the possibility of any shared function in the non-RPGs decreasing in splicing efficiency, I performed a Venn Diagram analysis to identify genes with

overlapping effects in Gcn5-HAT mutant background (Fig. 1G) followed by a GO analysis. There were no significant GO terms from component, process or function ontology with p-value cutoff of 0.05 that was found for the non-RPGs decreasing in SE under both *gcn5Δ* and *H3Δ9-16* [data not shown]. This data suggests that the Gcn5-dependent effects on non-RPG splicing outcomes are not determined by a shared cellular function, the associated wildtype RNA expression level or capacity for efficient splicing efficiency.

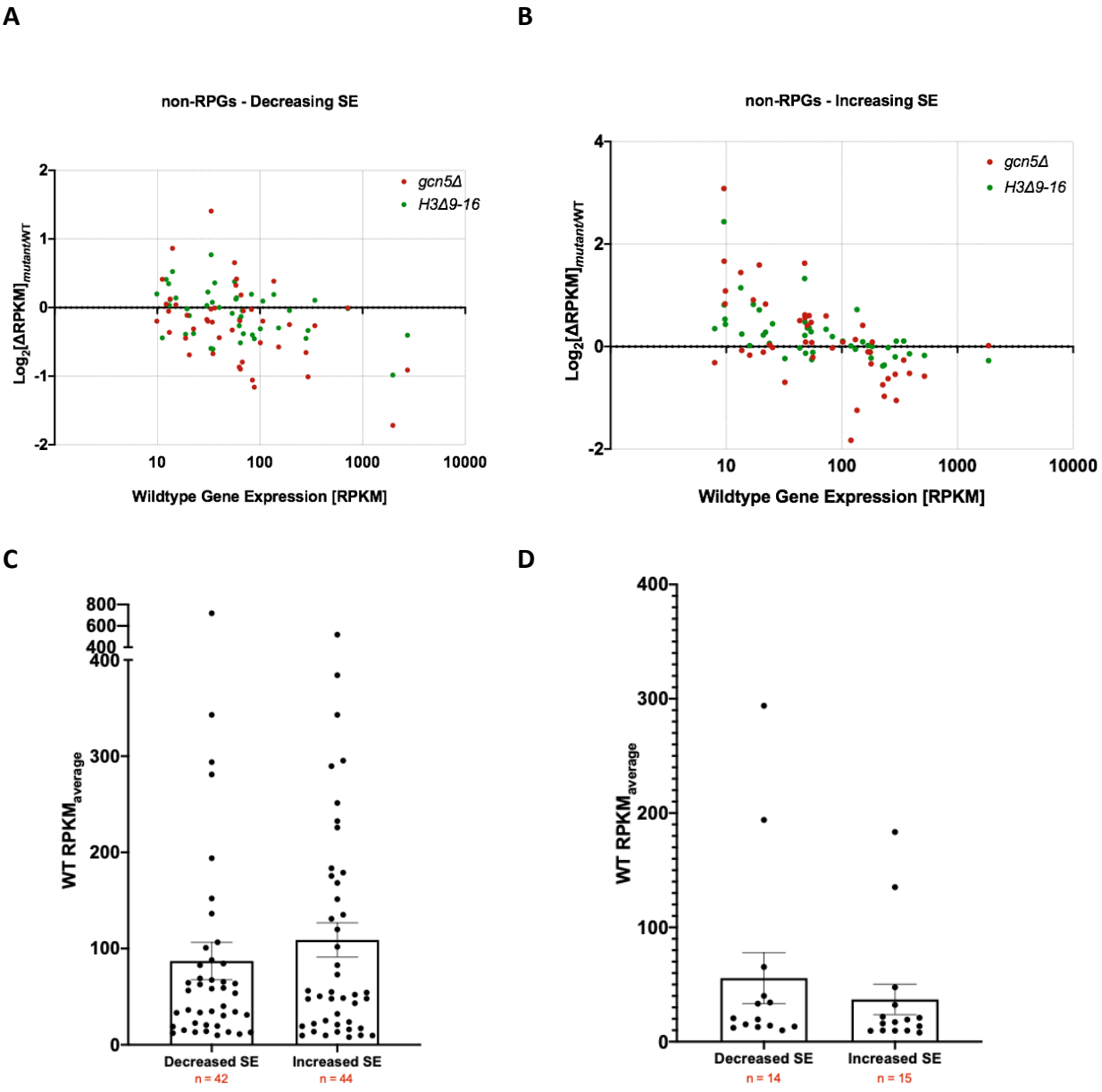


Figure 1 (A-D) - Wildtype RNA expression level and splicing efficiency does not determine Gcn5-HAT dependent splicing outcomes of IC-non-RPGs: A-B) Plot representing fold change in average RPKM for each IC-non-RPG compared to wildtype RPKM level based on A) decreasing splicing efficiency (SE) under mutants or B) increasing splicing efficiency under mutants. Mutant RPKM relative to wildtype RPKM after filtering (n = 127 total; spliced and unspliced read count ≥5 in WT). Each dot represents an IC-non-RPG represented as *gcn5 Δ* highlighted in red and *H3 Δ 9-16* are highlighted in green. Y-axis is the Log_2 transformation of the ratio of change in average RPKM level between either *gcn5 Δ* (P-value < 0.01) or *H3 Δ 9-16* (P-value < 0.01) and wildtype and x-axis is average WT RPKM level. C-D) Average WT RPKM level for C) filtered IC-non-RPGs with any decreased SE (n = 42) or increased SE (n = 44) under both mutants and D) filtered IC-non-RPGs with ≥5% decrease (n = 14) or increase (n = 15) in SE under both mutants. Error bars represent SEM of two biological replicate measurements of WT RPKM.

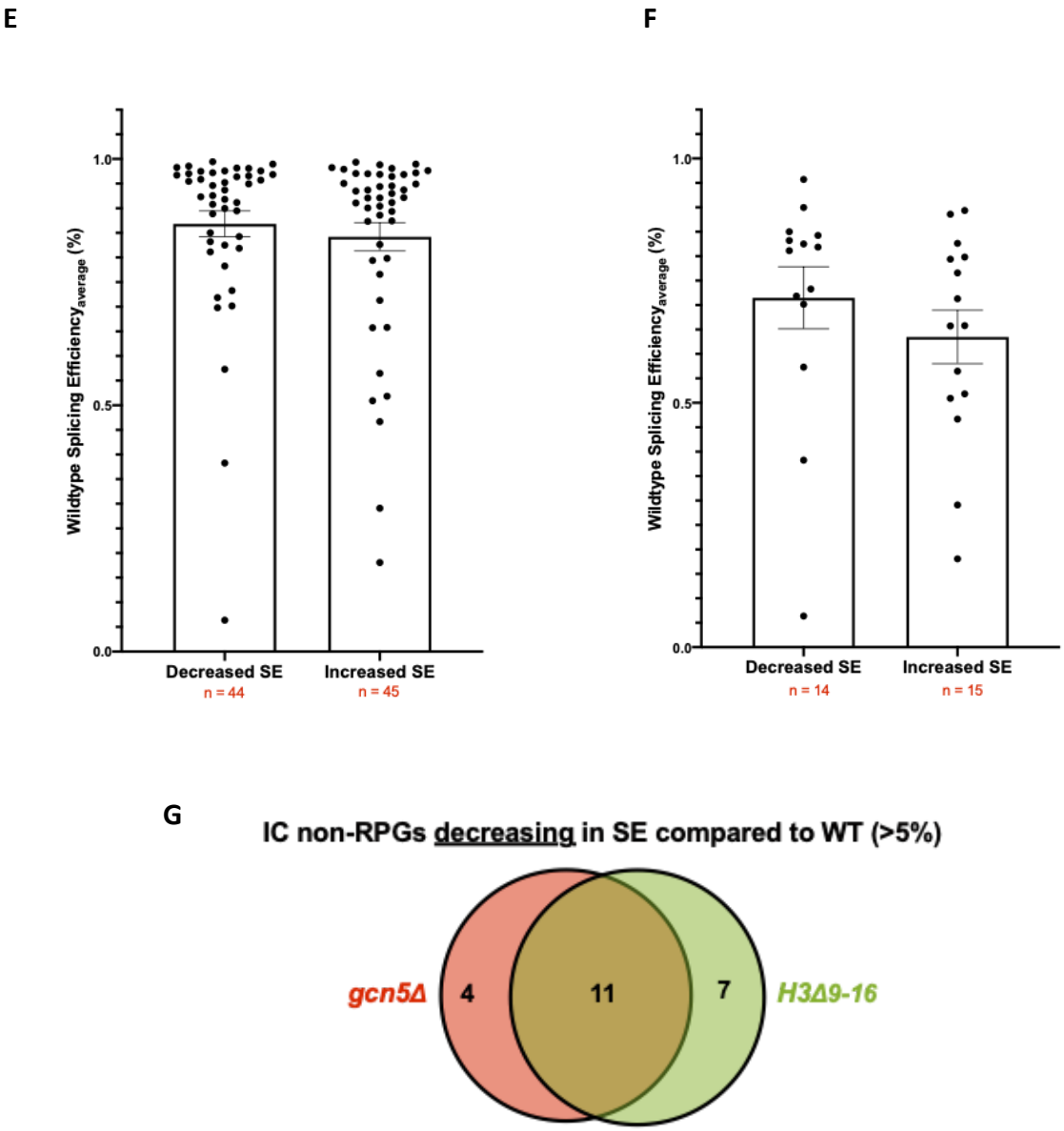


Figure 1 (E-G) - Wildtype RNA expression level and splicing efficiency does not determine Gcn5-HAT dependent splicing outcomes of IC-non-RPGs: E-F) Plot representing average % WT Splicing Efficiency for E) all filtered IC-non-RPGs with either any decreased SE (n = 44) or any increased SE (n = 45) under both mutants and F) all filtered IC-non-RPGs with ≥5% decrease (n = 14) or increase (n = 15) in SE under both mutants. Error bars represent SEM of two biological replicates. G) Venn Diagram representing IC-non-RPGs decreasing in SE compared to WT with at least ≥5% decrease in SE for each mutant.

***gcn5 Δ* and *H3 Δ9-16* dependent splicing outcomes of IC-non-RPGs are associated with distinct differences in RNAPII and H3K9ac profiles**

Characterizing chromatin architecture at genes is important for understanding the complex coordination of transcription and chromatin dynamics (Boeger et al., 2003; Carey et al., 2006; Carrozza et al., 2005; Henikoff, 2016; Joshi and Struhl, 2005). Given the co-transcriptional nature of splicing and since WT RNA expression level, splicing efficiency and shared cellular function have not determined *gcn5 Δ* and *H3 Δ9-16*-dependent splicing outcomes of IC-non-RPGs, I aimed to observe if there are inherent differences in the structure of chromatin at these groups of genes. If there are differences in the chromatin environment profiles correlated with *gcn5 Δ* and *H3 Δ9-16* dependent splicing outcomes, there may be a supporting evidence for the direct role of *gcn5 Δ* and *H3 Δ9-16* activity in co-transcriptional splicing.

To explore the contribution of chromatin structure to Gcn5-dependent splicing outcomes, I analyzed multiple ChIP-seq and MNase-seq datasets from Bruzzone et al. 2018 that assayed for the enrichment or recruitment of RNAPII, H3, H3K9ac and MNase sensitivity in a Gcn5+/- background (via rapamycin induced anchor-away method). My protocol was to categorize all of the non-RPGs by whether they decrease or increase in splicing efficiency in the *gcn5 Δ* and *H3 Δ9-16* mutants (overall and by at least 5% in my RNA-seq dataset), then analyze the average enrichment profile for each category via the chromatin signatures mentioned above. If there were differences in the enrichment profile of any particular category based on whether they decrease or increase in Gcn5-dependent splicing, then the chromatin landscape plays a role in determining splicing outcomes in a *gcn5 Δ* and *H3 Δ9-16* dependent context. Additionally,

consensus patterns between multiple chromatin signatures for any category of genes may hint at a pattern that validates biological significance of the categories (i.e. if one category shares both increased average RNAPII and increased average H3K9ac in similar loci, this may implicate increased chromatin accessibility for the category, on average). This data would begin to elucidate if the chromatin environment of non-RPGs decreasing in splicing efficiency under *gcn5Δ* and *H3Δ9-16* is the distinguishing reason for why they seem to be less sensitive to the RPG effect.

First, I analyzed the average recruitment profile of RNAPII for each category of genes. If the genes that are decreasing in splicing efficiency under deletion of *GCN5* have a reduced abundance of RNAPII at their loci, this may implicate a shared pattern of the binding state of RNAPII. Non-RPGs with decreased splicing efficiency ("SE down") contain on average an increased occupancy of RNAPII at the transcription start site (TSS) when compared to the non-RPGs increasing in splicing efficiency ("SE up") (Fig. 2A). This result is also observed in genes with greater than 5% change in splicing efficiency (Fig. 2B). Additionally, the occupancy profile of RNAPII downstream of the TSS varies distinctly between both categories--where non-RPGs decreasing in splicing efficiency demonstrate a steadily decreasing occupancy relative to the TSS and those that are increasing maintain a consistent level of occupancy throughout the gene. This data also demonstrates a decline and recovery of RNAPII near what seems to be the intronic region of the gene locus for those non-RPGs that are increasing in splicing efficiency. The same is not seen for the SE down category. Since RNAPII-Ser5p (the elongating form of post-translationally modified RNA Polymerase II) was immunoprecipitated in this

ChIP-seq experiment, this result has strengthened implications for the state of the enzyme at these loci. The elongating RNAPII at the non-RPGs decreasing in splicing efficiency may be losing affinity for the template DNA as transcription progresses--potentially hinting at physical chromatin barriers to processivity for this group of genes that are not as present for SE up. Another potential hypothesis for this result can be associated with each category of genes expressing differences in average elongation rate given that the average WT RNA transcript levels are consistent between categories. There may also be differences in the stalled/paused state of RNAPII between these categories. Without further experiments, it will be difficult to resolve the explicit state and rate of the RNAPII elongation complex in these categories of genes from occupancy and distribution alone.

To determine if *gcn5Δ* and *H3Δ9-16* mutation has a different degree of effect on the change in RNA expression level based on its splicing outcome in non-RPGs, I pooled the average fold change in RPKM level under both *GCN5* and the H3 mutant based on splicing outcome (>5% change) in *gcn5Δ* and *H3Δ9-16* mutation (Fig. 2C). I found that non-RPGs increasing in splicing efficiency under *gcn5Δ* and *H3Δ9-16*, tend to be increasing in RNA expression in both mutants as well. The non-RPGs decreasing in SE did not share this pattern in change, suggesting the functional state of RNAPII at these genes may be distinct from those in SE up. Collectively, this data reports an association between *gcn5Δ* and *H3Δ9-16* dependent splicing outcomes of non-RPGs and the locus-wide binding pattern of elongating RNAPII that is not due to WT RNA transcript level.

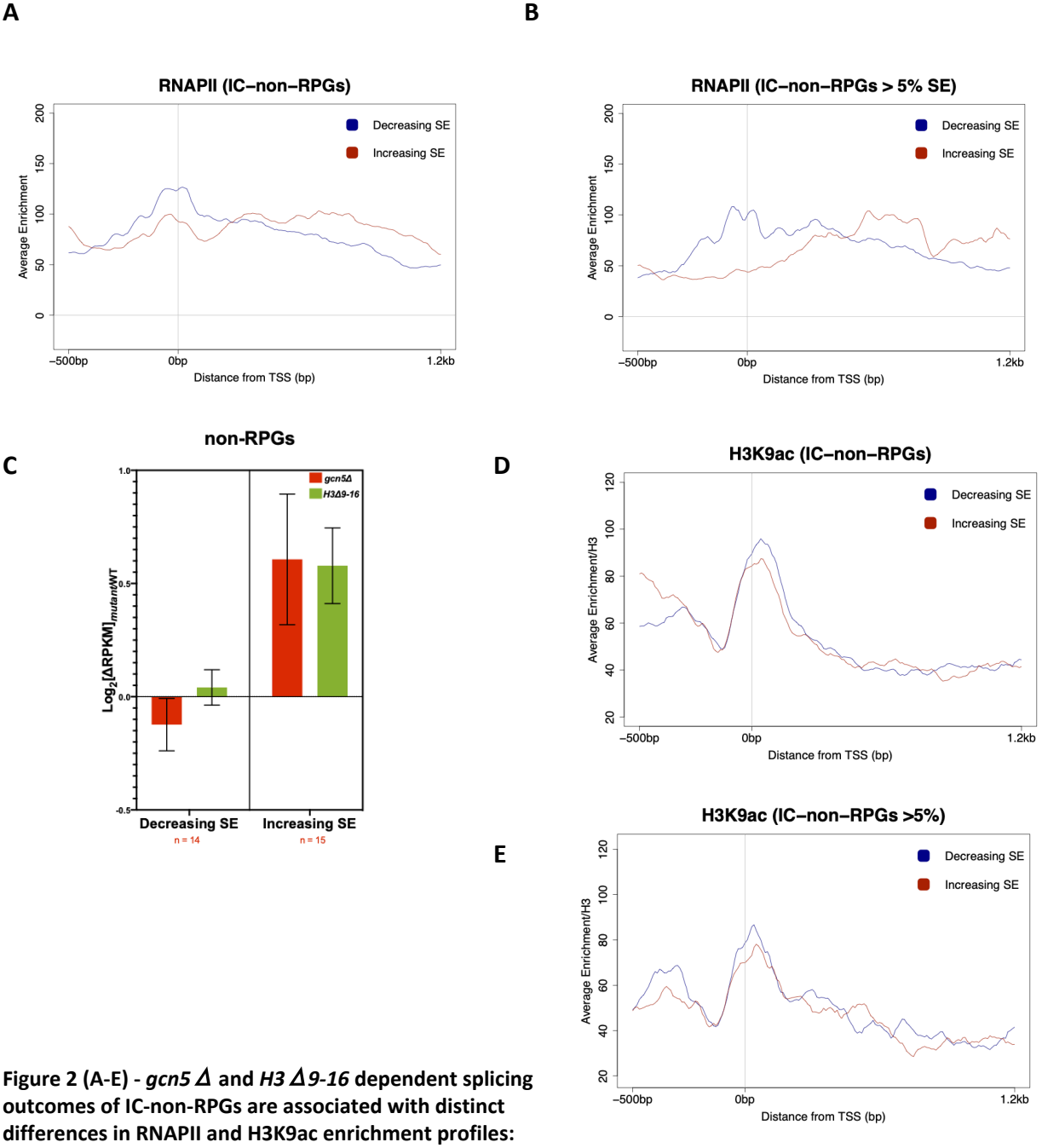
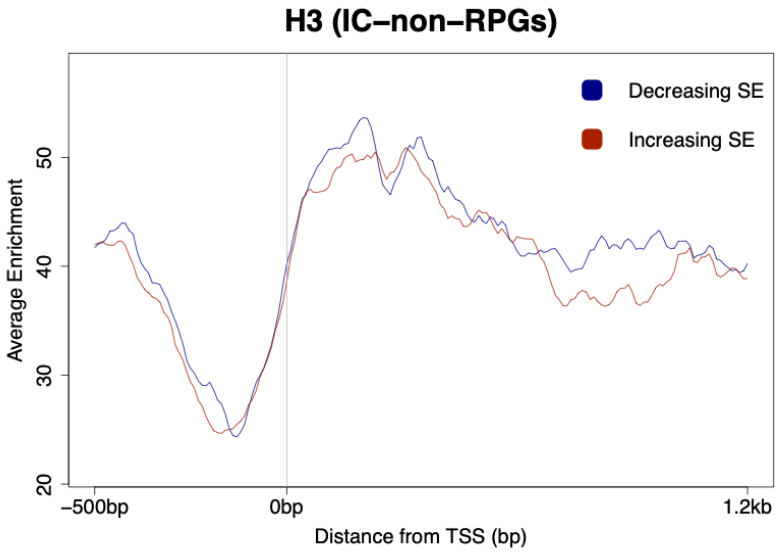


Figure 2 (A-E) - *gcn5 Δ* and *H3 Δ9-16* dependent splicing outcomes of IC-non-RPGs are associated with distinct differences in RNAPII and H3K9ac enrichment profiles: A-B) Meta-gene plot representing average enrichment profile of RNAPII-Ser5p for A) all filtered IC-non-RPGs either decreasing (n = 44, blue) or increasing (n = 45, red) in splicing efficiency by any amount and B) all filtered IC-non-RPGs with ≥5% decrease (n = 14) or increase (n = 15) in SE under both mutants. Y-axis represents the average enrichment per locus while x-axis represents the distance from the TSS. C) Plot representing fold change in average RPKM for each mutant relative to wildtype for non-RPGs either decreasing (n = 14) or increasing (n = 15) in SE. *gcn5 Δ* is highlighted in red and *H3 Δ9-16* i highlighted in green. Y-axis is the Log_2 transformation of the ratio of change in average RPKM level between either *gcn5 Δ* or *H3 Δ9-16* and wildtype. D-E) Meta-gene plot representing average enrichment profile of H3K9ac for D) all filtered IC-non-RPGs either decreasing (n = 44, blue) or increasing (n = 45, red) in splicing efficiency by any amount and E) all filtered IC-non-RPGs with ≥5% decrease (n = 14) or increase (n = 15) in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.

F



G

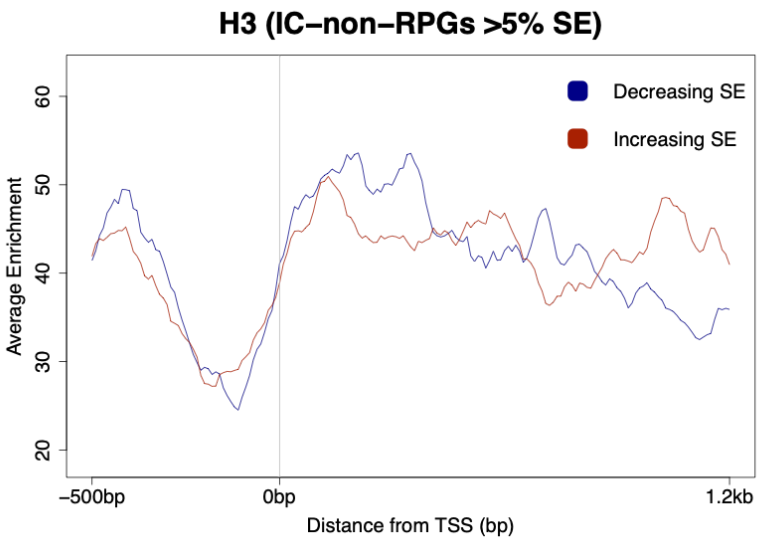


Figure 2 (F-G) - *gcn5 Δ* and *H3 Δ9-16* dependent splicing outcomes of IC-non-RPGs are associated with distinct differences in RNAPII and H3K9ac enrichment profiles: F-G) Meta-gene plot representing average enrichment profile of H3 for F) all filtered IC-non-RPGs either decreasing (n = 44, blue) or increasing (n = 45, red) in splicing efficiency by any amount and G) all filtered IC-non-RPGs with ≥5% decrease (n = 14) or increase (n = 15) in SE under both mutants. Y-axis represents the average enrichment per locus while x-axis represents the distance from the TSS.

I next asked whether there were differences in the average H3K9ac profile across the loci between each category of genes. The enrichment of H3K9ac does not differ in the main gene body based on splicing outcomes (Fig. 2D). However, for non-RPGs decreasing in splicing efficiency, there is an increased peak of H3K9ac near the TSS relative to the profile of non-RPGs increasing in splicing efficiency (Fig. 2D and 2E). As Gcn5p is the major yeast histone acetyltransferase, the enriched TSS-proximal H3K9ac at the non-RPGs decreasing in splicing efficiency suggests that this category of genes may be more sensitive to a lack of *GCN5* and H3K9-16 acetylation activity and that this is sensitivity is potentially resulting in the decreased splicing splicing efficiency. This result is intriguing as previous reports have shown that histone acetylation by Gcn5 stimulates nucleosome binding and remodeling by the SWI/SNF complex (Chandy et al. 2006; Chatterjee et al. 2011). To determine if the increased H3K9ac was a product of increased H3 at the loci for these genes, I analyzed H3 enrichment across the gene body for both categories. While Gcn5p depletion resulted in decreased H3 across the gene in both groups, there was no marked difference in WT abundance of H3 across the gene in either group (Fig. 2F and 2G).

gcn5Δ and H3Δ9-16 dependent splicing outcomes of IC-non-RPGs are associated with differences in nucleosome positioning

In recent years, nucleosomes have been implicated in having a role in splicing. For example, Chen and colleagues observed higher occupancy of nucleosomes arranged at exonic sequences whereas intron sequences displayed lower occupancy in mammals (Chen et al., 2010). To determine if there was a difference in the nucleosome occupancy profile of non-RPGs based on

Gcn5-dependent effects on splicing, I plotted MNase-seq distributions for each category (Fig. 3). Intriguingly, I found distinct distribution patterns of nucleosome occupancy associated with splicing outcomes in Gcn5-HAT mutant background. More specifically, non-RPGs with decreasing splicing efficiency demonstrate a well-defined MNase-seq digestion array across the gene whereas those that are increasing in splicing efficiency tend to display fuzziness in the gene body. Additionally, the height of the nucleosome occupancy peaks are distinct between groups. The SE down category of non-RPGs is associated with loci that are strongly sensitive to MNase digestion, while the SE up category represents otherwise. Given the chromatin implications of MNase sensitivity at gene loci, the different profiles between non-RPGs decreasing and increasing in splicing efficiency reflect a difference in the chromatin environment associated with Gcn5-dependent splicing outcomes.

For SE down, it appears as though the well-defined MNase-seq digestion signature reflects nucleosomes that are more stably associated with the DNA. In contrast, peaks with more modest amplitude and less clear dips (as represented by the non-RPGs increasing in splicing efficiency) demonstrate a signature associated with more mobile nucleosomes. Given the results from Figure 2B, SE up IC-non-RPGs are presumably more active and therefore the nucleosomes are more likely to be sliding around relative to genes that are being down-regulated which could have more stable nucleosomes. Since metagene plots are an average signal distribution over a group of genes, nuanced information about the chromatin environment on a single-gene basis may be lost. To hone in on the the individual chromatin environment IC-non-RPGs with >5% SE in either direction, I repeated the above analysis on

individual IC-non-RPGs (Supplementary Figures). I also provide representative IGV tracks for each gene (Supplementary Figures). Together with the above data, IC-non-RPGs that are decreasing in splicing efficiency under Gcn5-dependent splicing may demonstrate a WT chromatin environment that is designed for less mobile gene regulatory elements.

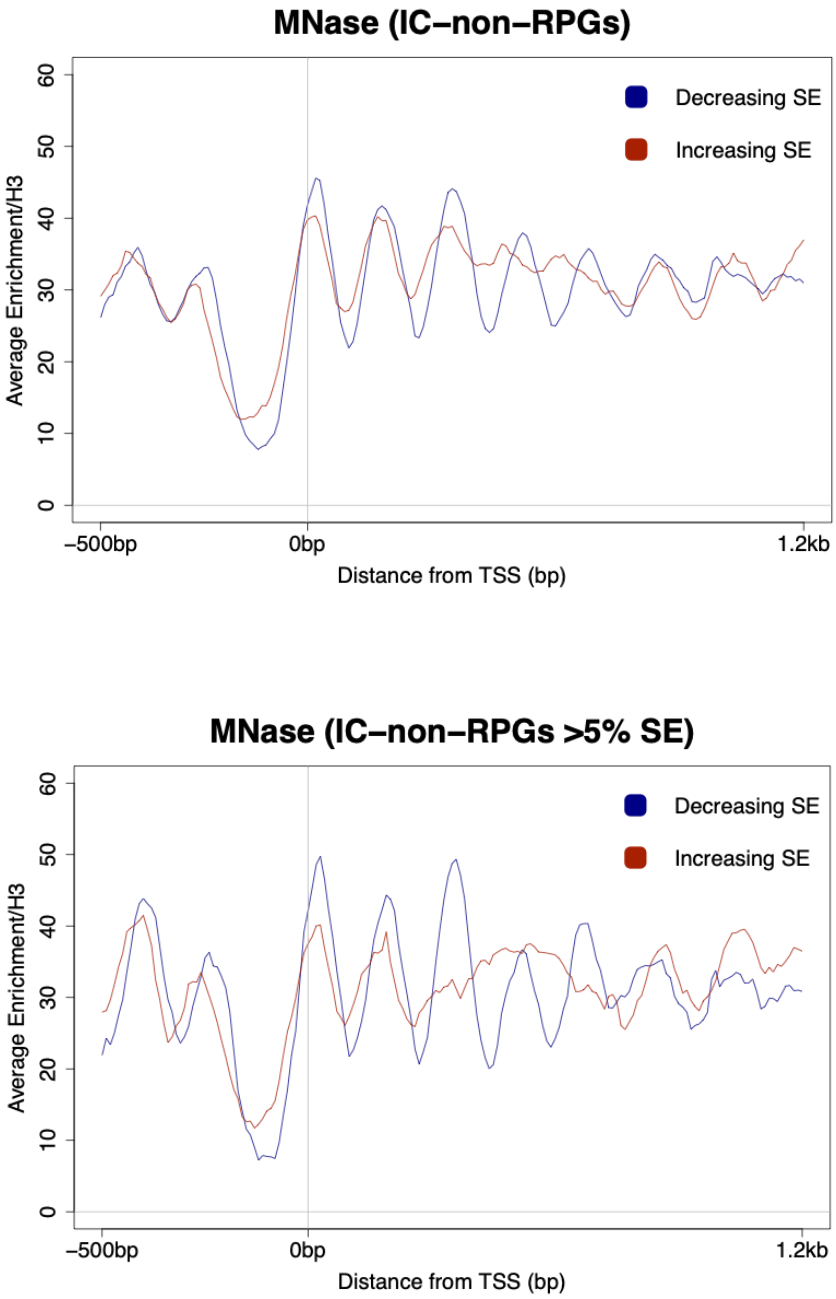


Figure 3 - *gcn5 Δ* and *H3 Δ9-16* dependent splicing outcomes of IC-non-RPGs are associated with differences in nucleosome positioning: A-B) Meta-gene plot representing average enrichment profile of MNase digestion for A) all filtered IC-non-RPGs either decreasing (n = 44, blue) or increasing (n = 45, red) in splicing efficiency by any amount and B) all filtered IC-non-RPGs with ≥5% decrease (n = 14) or increase (n = 15) in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.

DISCUSSION

My data provides support for the state of chromatin architecture determining Gcn5-dependent splicing outcomes for non-RPGs (Fig. 4). This observation also provides support for non-RPGs decreasing in *gcn5 Δ* and *H3 Δ 9-16* dependent splicing being less sensitive to the increased availability of spliceosomes due to the RPG effect. The result that *gcn5 Δ* and *H3 Δ 9-16* dependent splicing outcomes are associated with differences in MNase sensitivity profiles raises several mechanistic hypotheses regarding the state of the nucleosome environment and its connection to splicing. Potentially, these observations may support the model for *gcn5 Δ* and *H3 Δ 9-16* function in facilitating direct recruitment of splicing factors to chromatin. Under this model, a stable nucleosome landscape provides the opportunity for Gcn5 protein or Gcn5-modified lysines on H3 to physically recruit splicing factors. *gcn5 Δ* and *H3 Δ 9-16* would decrease recruitment of splicing factors to chromatin and result in decreased splicing outcomes. Alternatively, the *gcn5 Δ* and *H3 Δ 9-16* mutation may be influencing transcription elongation rate such to sensitize groups of genes with a particular chromatin environment to collaborate less effectively with co-transcriptional spliceosome assembly.

Both of these models support a role for *gcn5 Δ* and *H3 Δ 9-16* dependent activity in pre-mRNA splicing genome-wide, within the additional context of the RPG effect on spliceosome distribution. The experimental analyses in Chapter 3 were performed to identify why the subset of IC-non-RPGs decreasing in SE seem to be less sensitive to the RPG effect. The nature of the chromatin landscape, particularly regarding the mobility of nucleosomes, may be involved in establishing precursory limitations for co-transcriptional splicing outcomes, despite

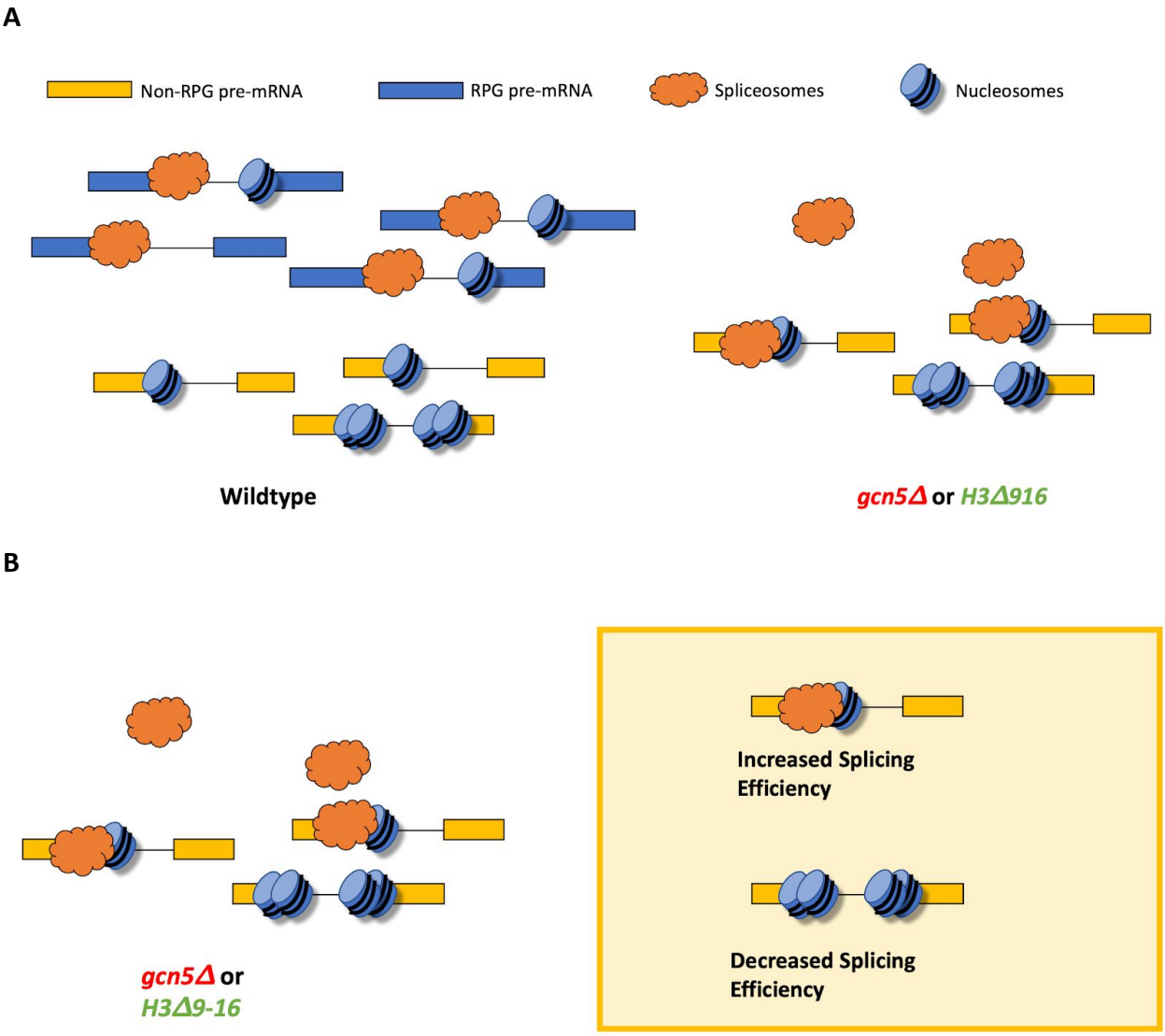


Figure 4 - Model for the relationship between Gcn5-histone acetyltransferase activity and pre-mRNA splicing

the availability of spliceosomes. Hypothetically speaking, if a chromatin environment contains a nucleosome occupancy profile that is less permissive to recruiting splicing factors or facilitating an optimized transcription elongation rate, the perceived benefit (for IC-non-RPGs) of an increased availability of spliceosomal factors will not not be realized due to the natural barriers of the chromatin at these gene loci.

RPGs are known to be the most highly-expressed and transcript abundant class of genes in yeast and share a chromatin landscape marked by fragile nucleosomes. For example, fragile nucleosomes are defined by multiple parameters such as relative sensitivity to MNase, stability and distance between the -1/+1 nucleosome, competitive binding of general transcription factors, sequence motifs, etc. RPGs may represent a more transcriptionally plastic class of genes--supported by increased nucleosome fragility and aptitude for environmental/nutrient stress response and etc. Recent reports implicated the occupancy of the RSC complex, a multi-subunit complex that binds to acetylated histones, as an identifier of fragile nucleosomes. The Rsc4 protein of the RSC complex is inactivated by Gcn5 acetylation at Lys 25 and as a result, Gcn5 demonstrates inhibited binding of its bromodomain to acetylated H3K14 (VanDemark et al., 2007). *gcn5Δ* and *H3Δ9-16* dependent activity may be functioning directly at these fragile nucleosomes for IC-RPGs and non-IC-RPGs, thus providing more support for the physical recruitment model of spliceosome assembly.

The MNase results are striking yet invite complex interpretations. One of the hypotheses I originally aimed to explore was whether IC-non-RPGs with decreased splicing efficiency under

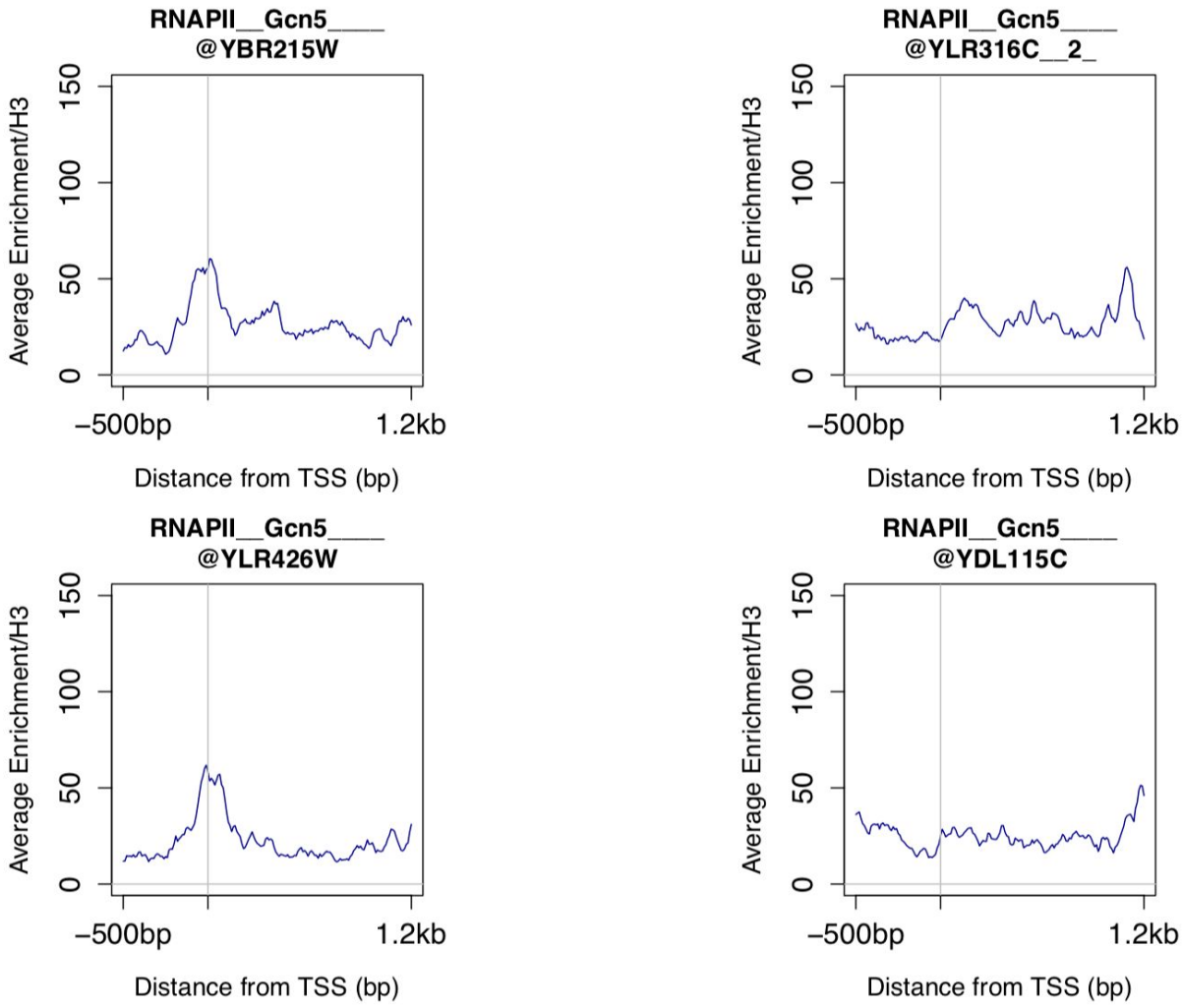
GCN5 have a MNase-seq profile reflective of nucleosomes that are less likely to be displaced which may potentially be more resilient to changes in spliceosome availability. This would be further supported if IC-non-RPGs with increased splicing efficiency demonstrate the inverse profile. If this group of genes harbor a chromatin environment that is primed for the type of regulation that does not require highly mobile regulatory elements or activity, the RPG effect induced availability of spliceosomes may meet physical barriers at these loci, specifically. Nucleosomes by definition are barriers for transcriptional machinery and can regulate access of these factors to the protected DNA template. Per the data I presented in this chapter, the IC-non-RPGs with decreased splicing efficiency may be less sensitive the RPG effect given a chromatin environment that is either more or less protected (or stabilized) by nucleosomes. This implicates a direct physical mechanism for the relationship between Gcn5 and spliceosome assembly.

In order to fully elucidate the specific mechanism that is underlying why a subset of non-RPGs are less sensitive to the availability of spliceosomes, further experiments are needed. For example, if the physical recruitment model is determining splicing outcomes in a *gcn5Δ* and *H3Δ9-16* dependent manner, identifying the localization of Gcn5 at these genes would be of great interest. If Gcn5 is enriched at these genes, it would strongly implicate its role in physically recruiting spliceosome components. This data would be strengthened by accompanying ChIP-seq of factors associated with sequential spliceosome components (U1 snRNP, U2 snRNP, prp19, etc). NET-seq in a *GCN5* mutant background would also help to

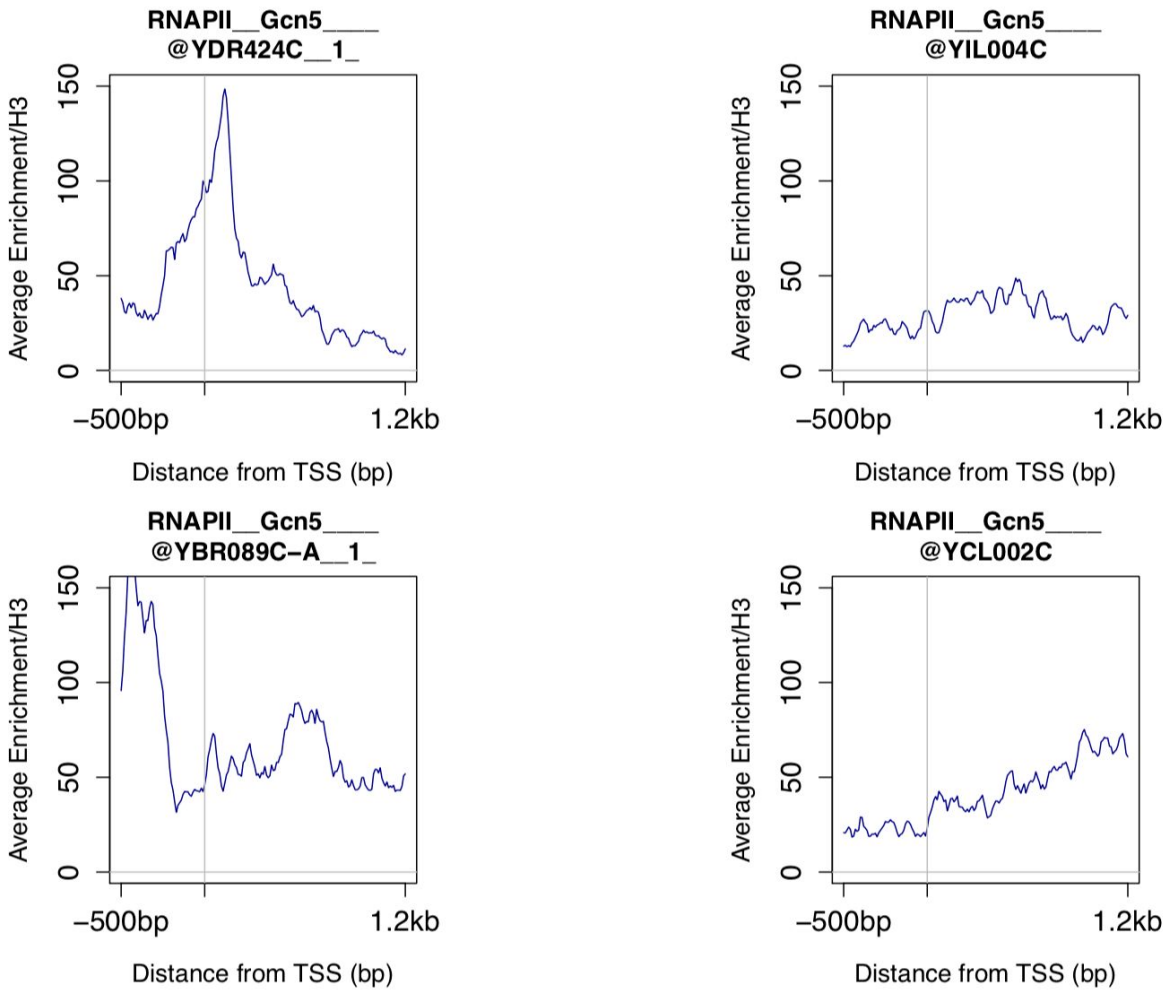
determine average elongation rates for each group of non-RPGs to discriminate between the physical and kinetic relationship of *gcn5 Δ* and *H3 Δ 9-16* to splicing outcomes.

Overall, this scientific body of work presents an unique methodology for understanding the varied effects of *gcn5 Δ* and *H3 Δ 9-16* in pre-mRNA splicing. Prior to this project, the genome-wide effects of *gcn5 Δ* and *H3 Δ 9-16* in splicing were not known and widely accepted understanding that Gcn5 does not affect global transcription, it may have been inferred that there is no role for *gcn5 Δ* and *H3 Δ 9-16* in global splicing outcomes. My data uncovered multiple effects of the *gcn5 Δ* and *H3 Δ 9-16* on pre-mRNA splicing genome-wide--encompassing mechanisms regarding the competition of limited spliceosomes and distinct chromatin landscapes directly regulating splicing outcomes. Given the highly regulated nature of Gcn5 histone acetylation, these results provide intriguing context for the collaboration between the chromatin environment and spliceosome assembly. Whether the relationship is predominantly governed by a kinetic relationship or a physical relationship will be of future interest. Altogether, it can be appreciated that Gcn5 plays a complex role in genome-wide pre-mRNA splicing that factors both the economics of spliceosome distribution and natural state of the chromatin.

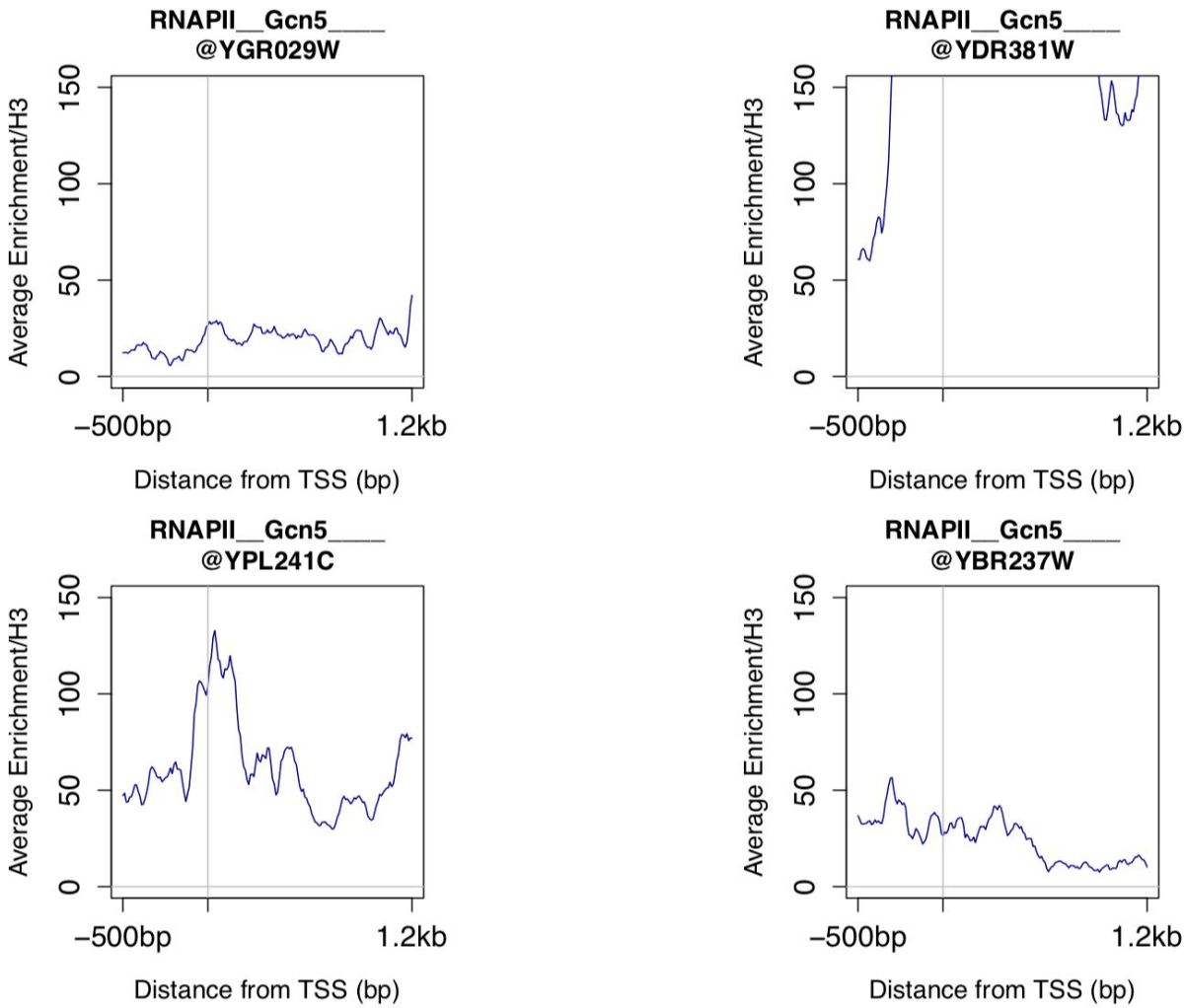
Supplementary Figures:



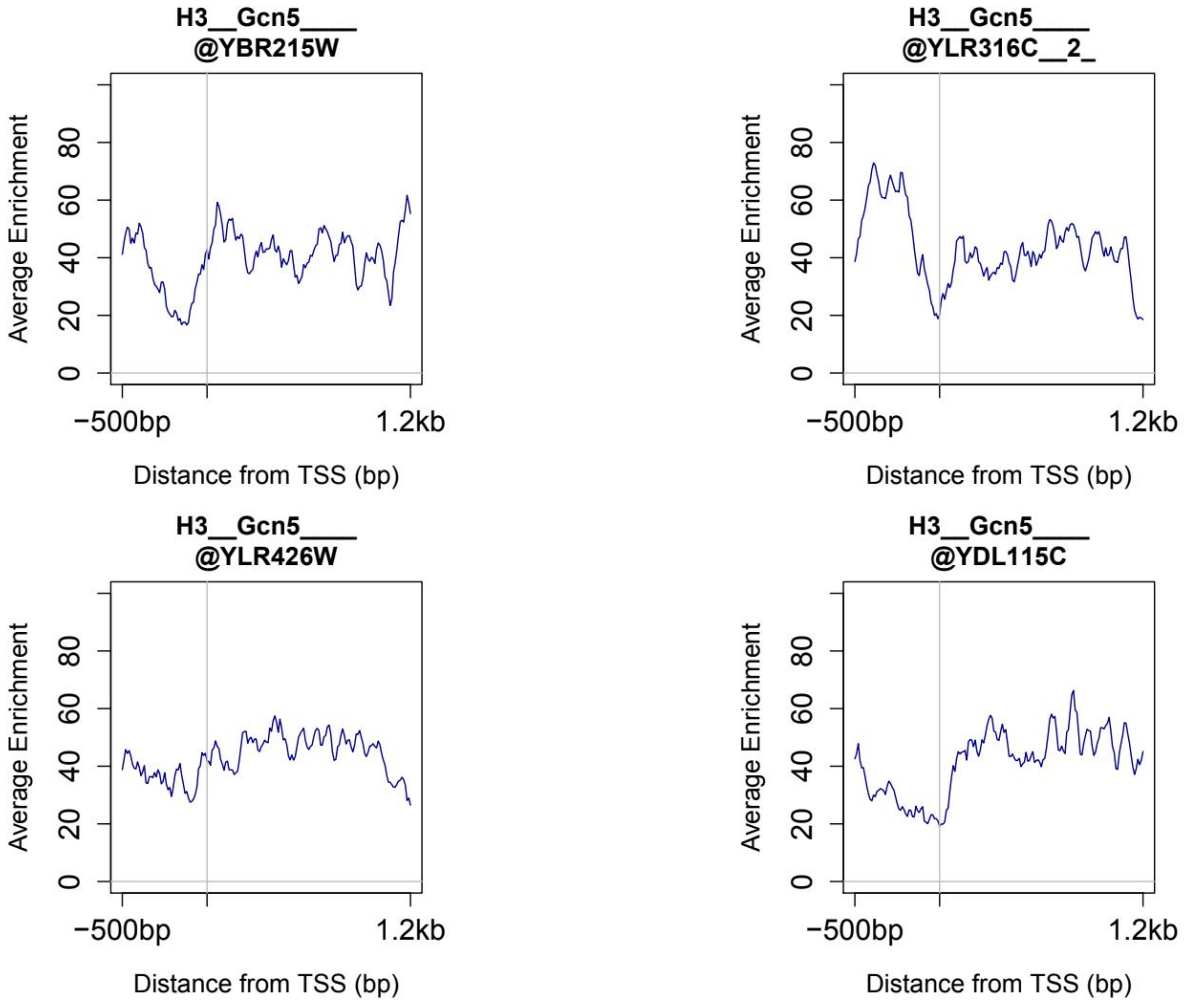
Supplementary Figure 1: Each meta-gene plot represents the average ChIP-seq enrichment profile of RNAPII-Ser5p for individual IC-non-RPGs demonstrating ≥5% decrease in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.



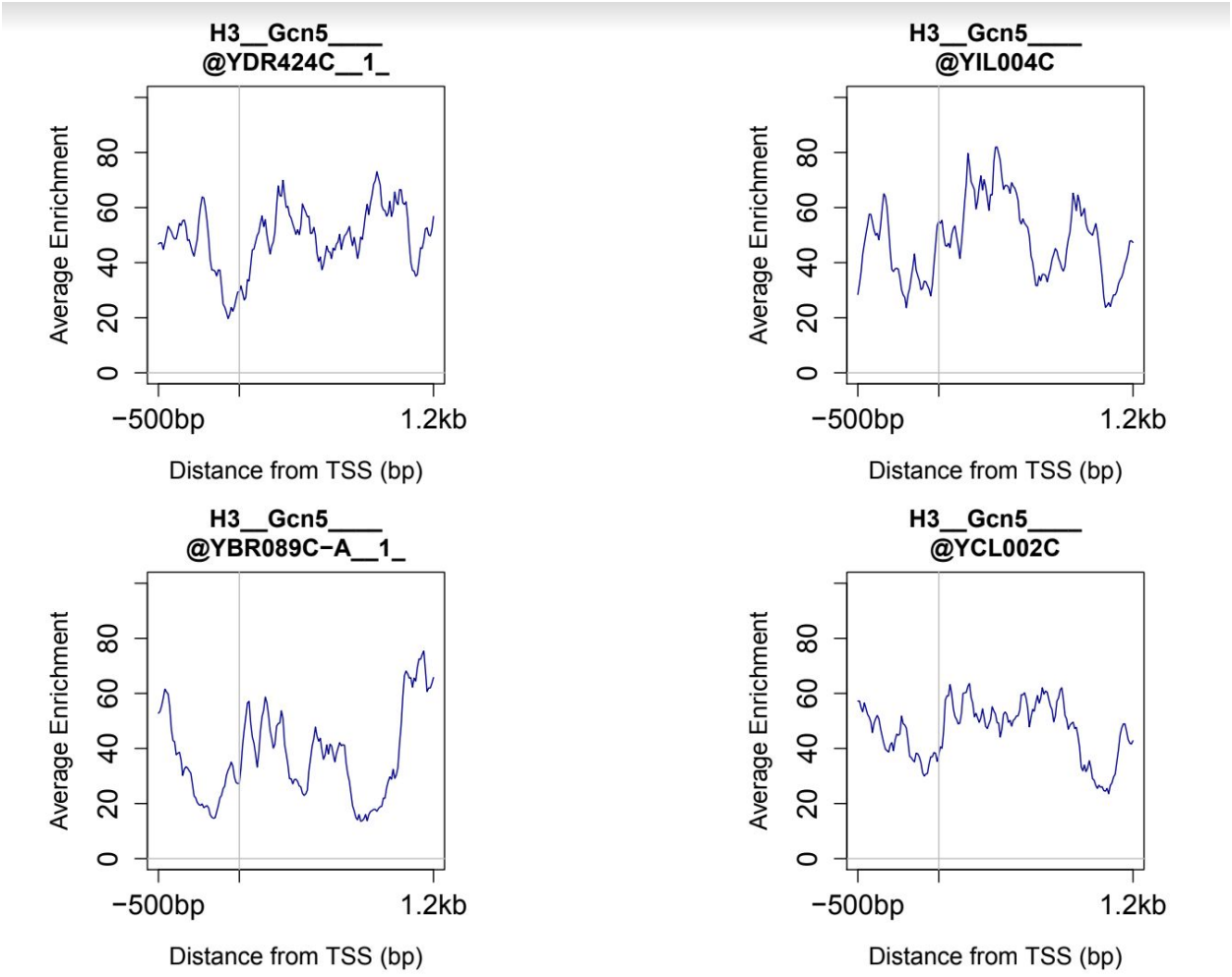
Supplementary Figure 1 (cont'd): Each meta-gene plot represents the average ChIP-seq enrichment profile of RNAPII-Ser5p for individual IC-non-RPGs demonstrating ≥5% decrease in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.



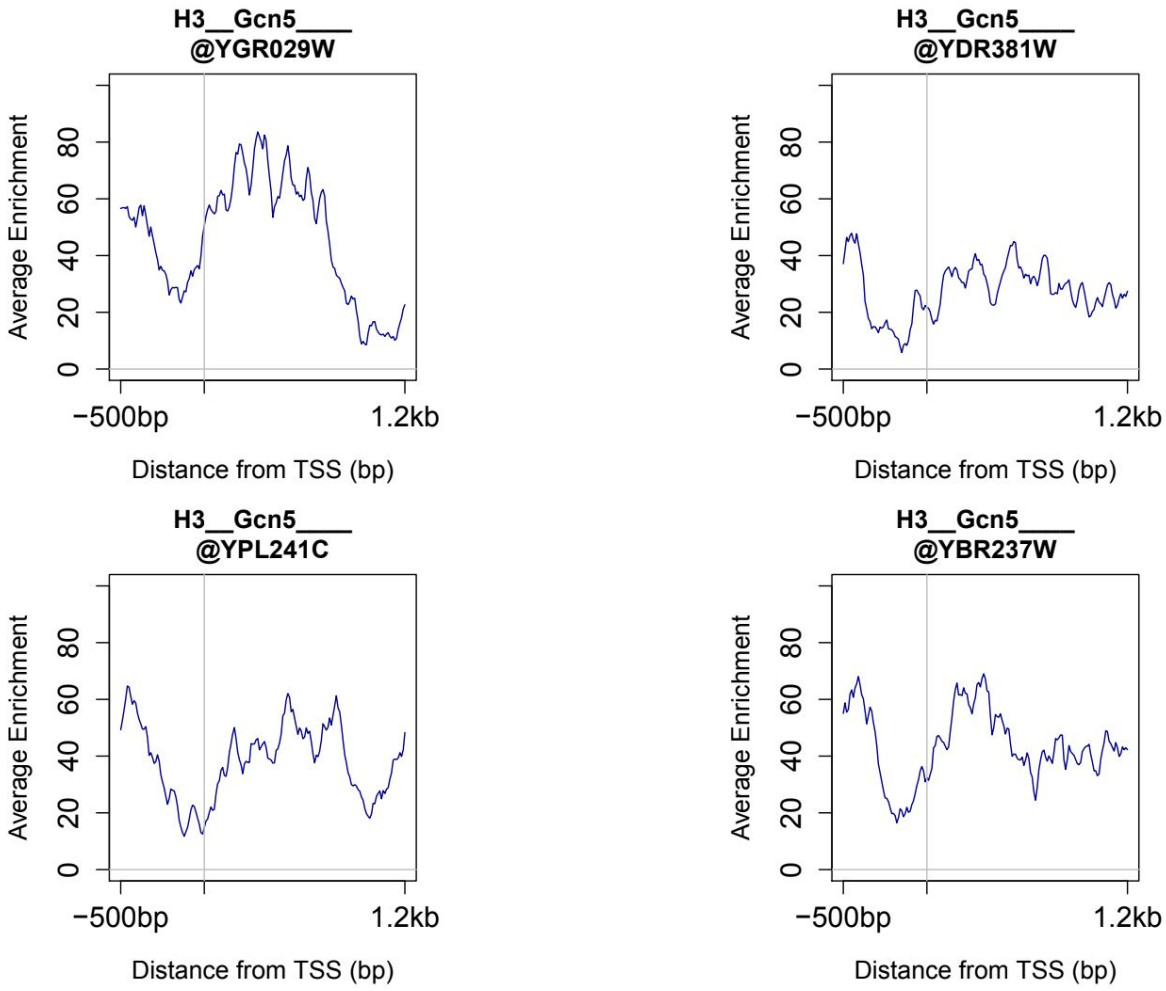
Supplementary Figure 1 (cont'd): Each meta-gene plot represents the average ChIP-seq enrichment profile of RNAPII-Ser5p for individual IC-non-RPGs demonstrating ≥5% decrease in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.



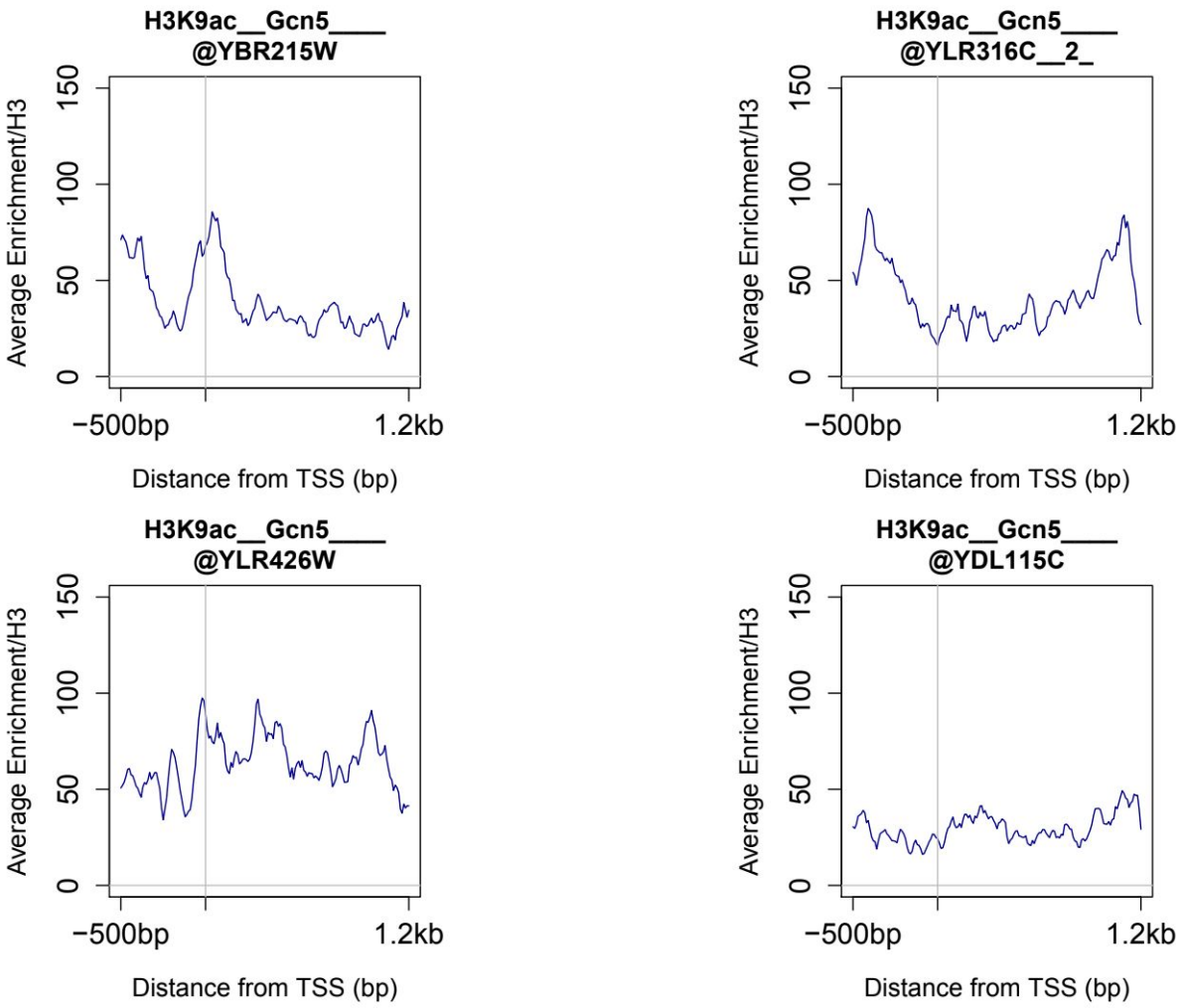
Supplementary Figure 2: Each meta-gene plot represents the average ChIP-seq enrichment profile of H3 for individual IC-non-RPGs demonstrating ≥5% decrease in SE under both mutants. Y-axis represents the average enrichment per locus while x-axis represents the distance from the TSS.



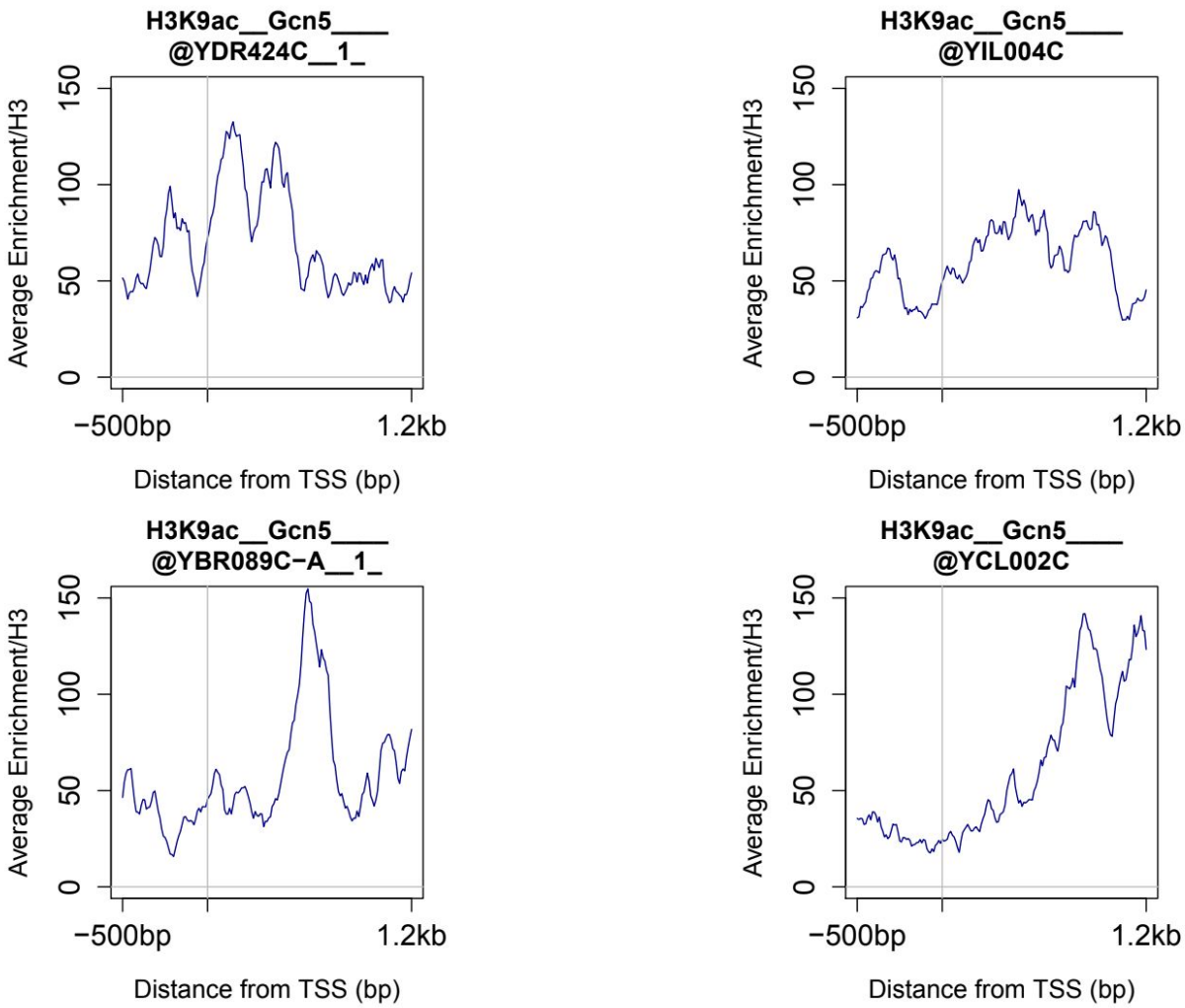
Supplementary Figure 2 (cont'd): Each meta-gene plot represents the average ChIP-seq enrichment profile of H3 for individual IC-non-RPGs demonstrating ≥5% decrease in SE under both mutants. Y-axis represents the average enrichment per locus while x-axis represents the distance from the TSS.



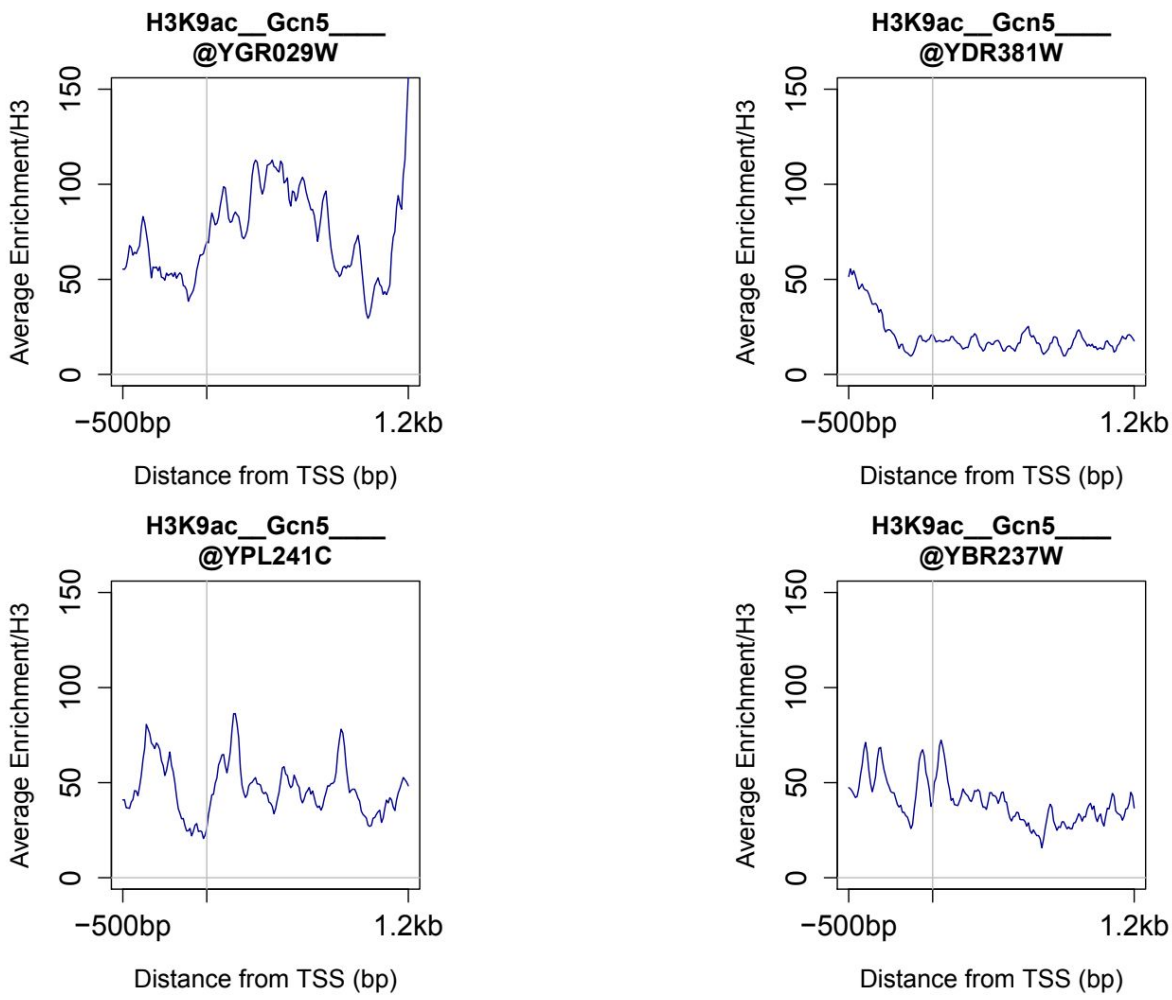
Supplementary Figure 2 (cont'd): Each meta-gene plot represents the average ChIP-seq enrichment profile of H3 for individual IC-non-RPGs demonstrating ≥5% decrease in SE under both mutants. Y-axis represents the average enrichment per locus while x-axis represents the distance from the TSS.



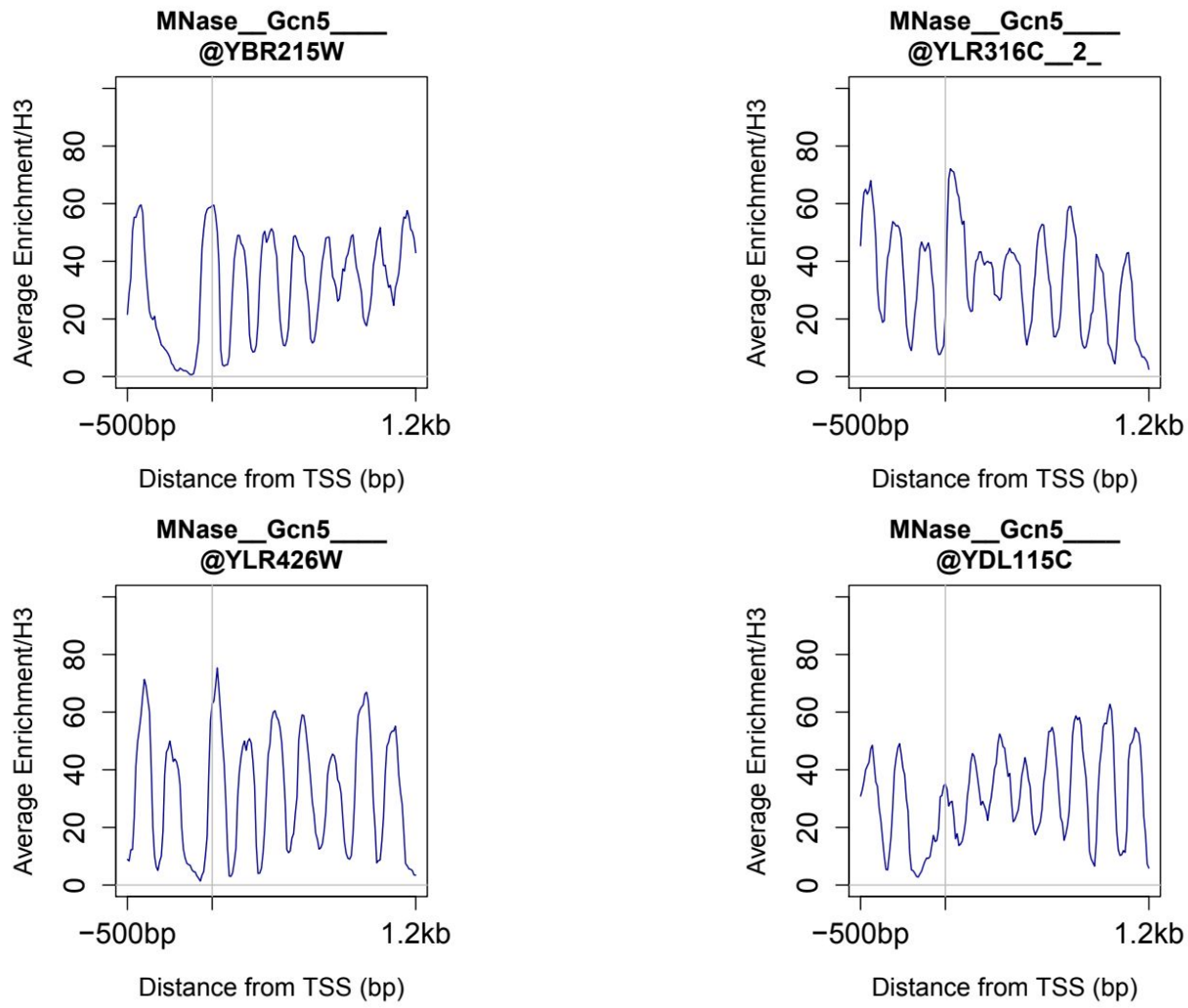
Supplementary Figure 3: Each meta-gene plot represents the average ChIP-seq enrichment profile of H3K9ac for individual IC-non-RPGs demonstrating ≥5% decrease in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.



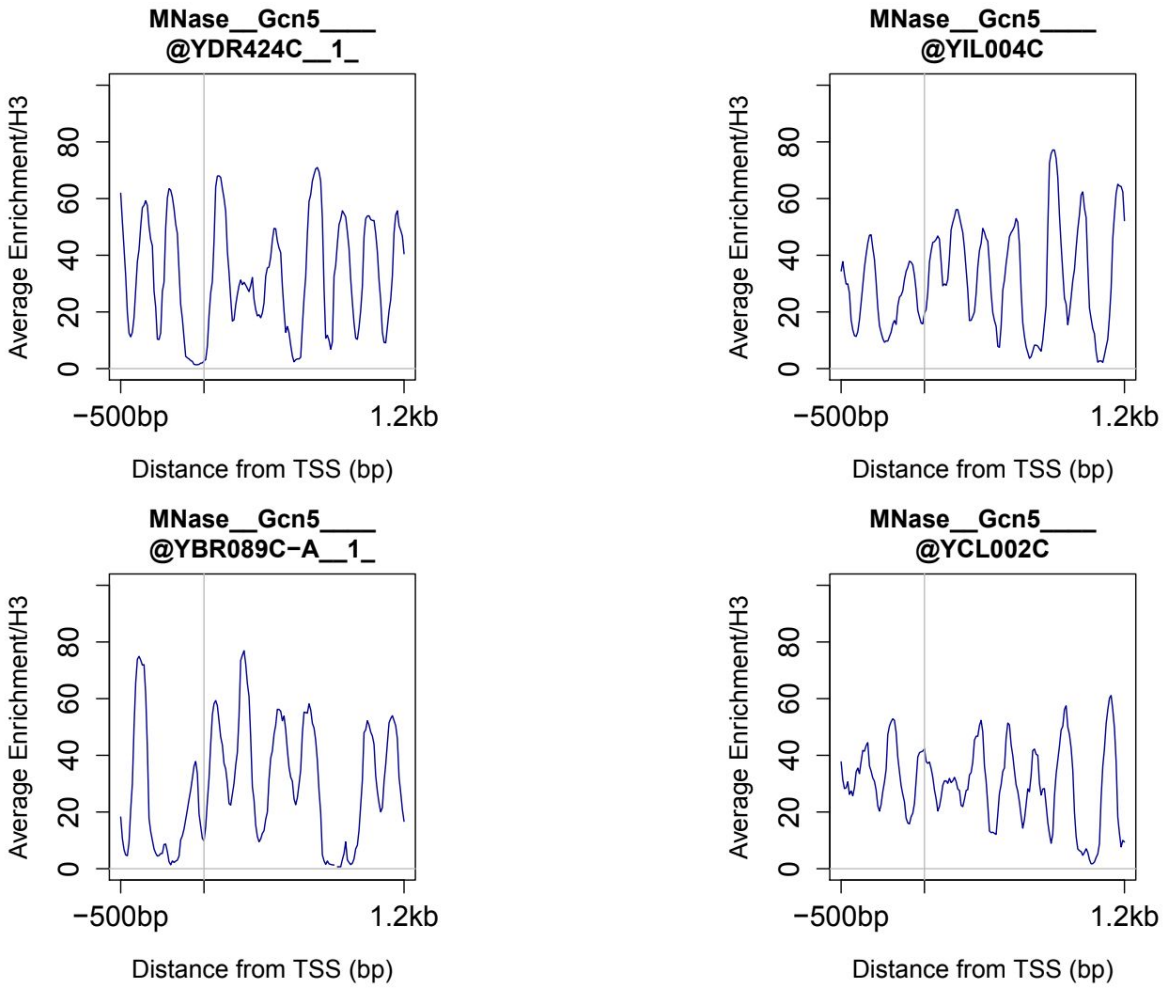
Supplementary Figure 3 (cont'd): Each meta-gene plot represents the average ChIP-seq enrichment profile of H3K9ac for individual IC-non-RPGs demonstrating ≥5% decrease in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.



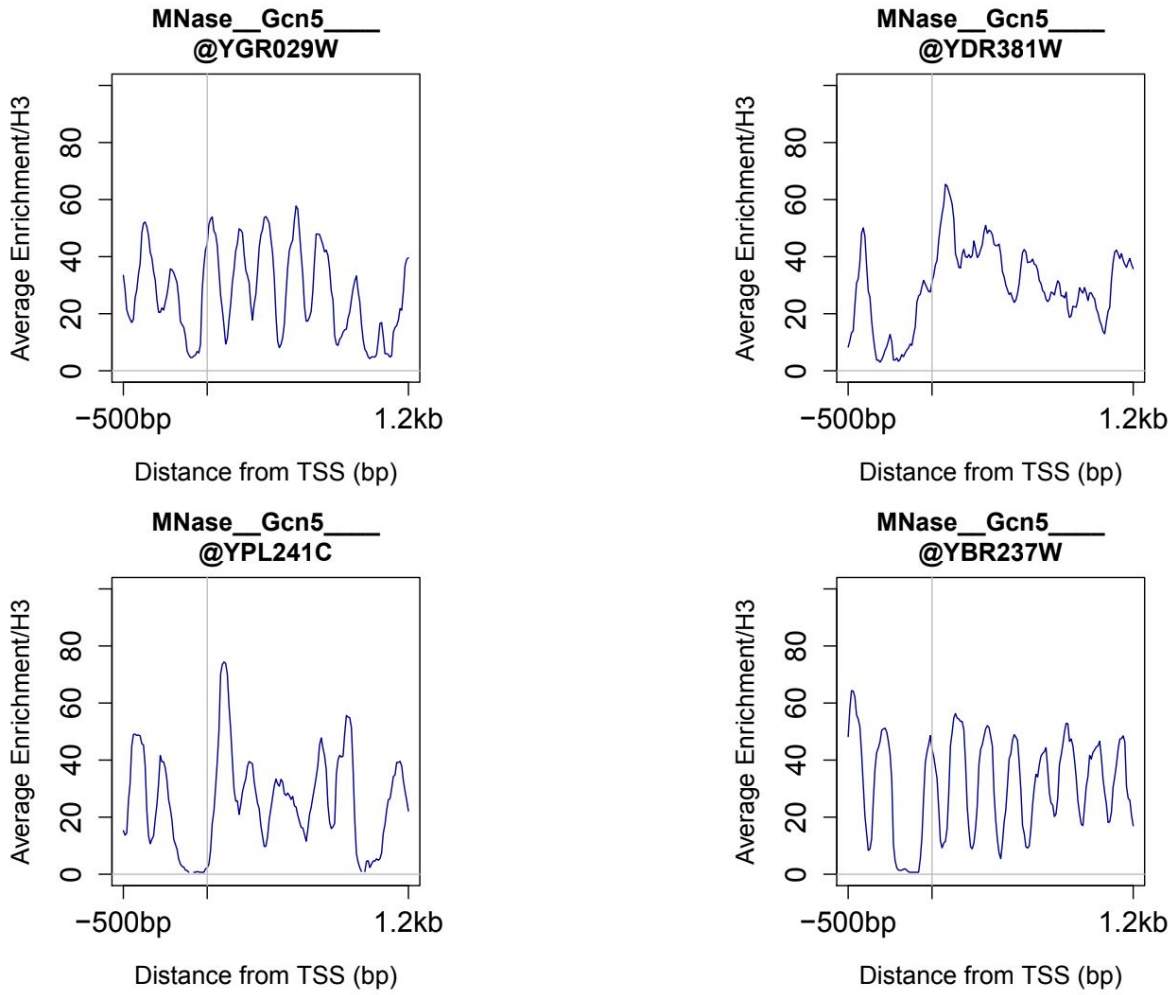
Supplementary Figure 3 (cont'd): Each meta-gene plot represents the average ChIP-seq enrichment profile of H3K9ac for individual IC-non-RPGs demonstrating ≥5% decrease in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.



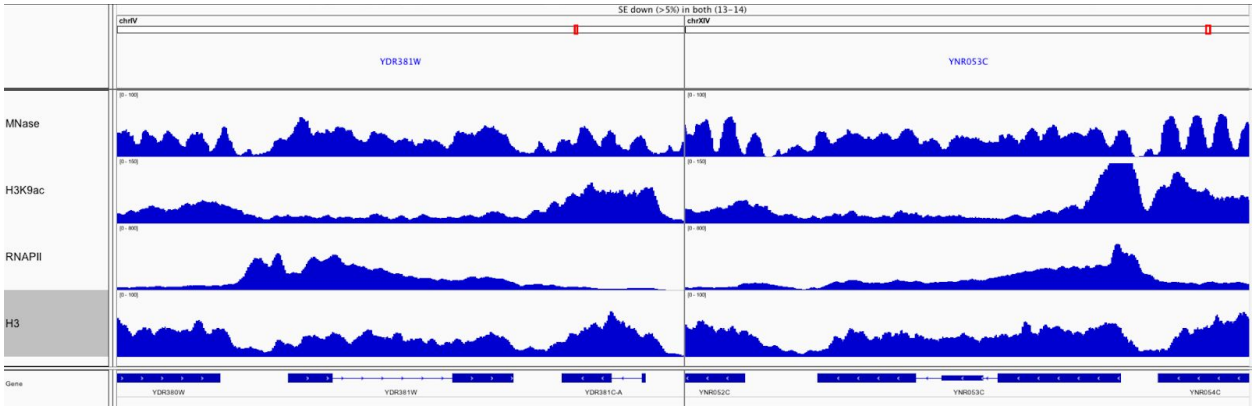
Supplementary Figure 4: Each meta-gene plot represents the average enrichment profile of MNase-seq digestion for individual IC-non-RPGs demonstrating ≥5% decrease in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.

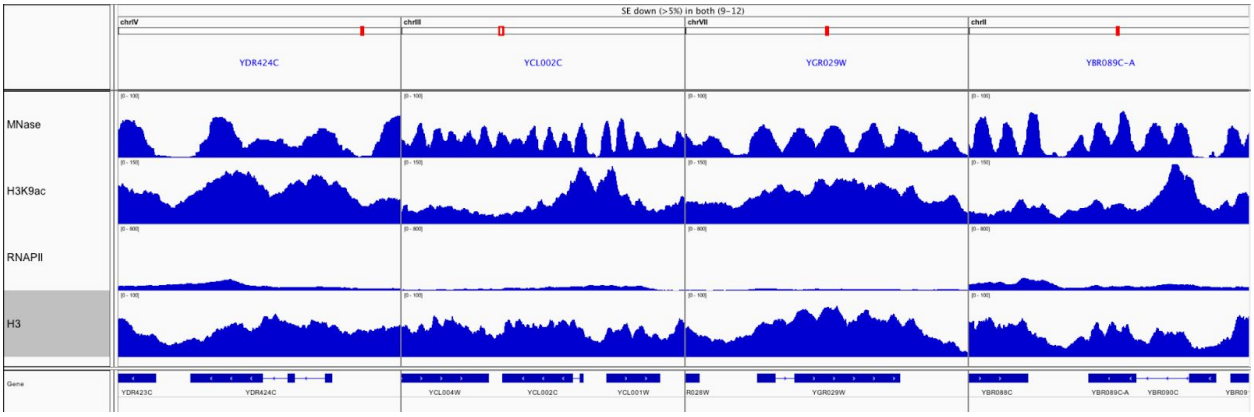


Supplementary Figure 4 (cont'd): Each meta-gene plot represents the average enrichment profile of MNase-seq digestion for individual IC-non-RPGs demonstrating ≥5% decrease in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.

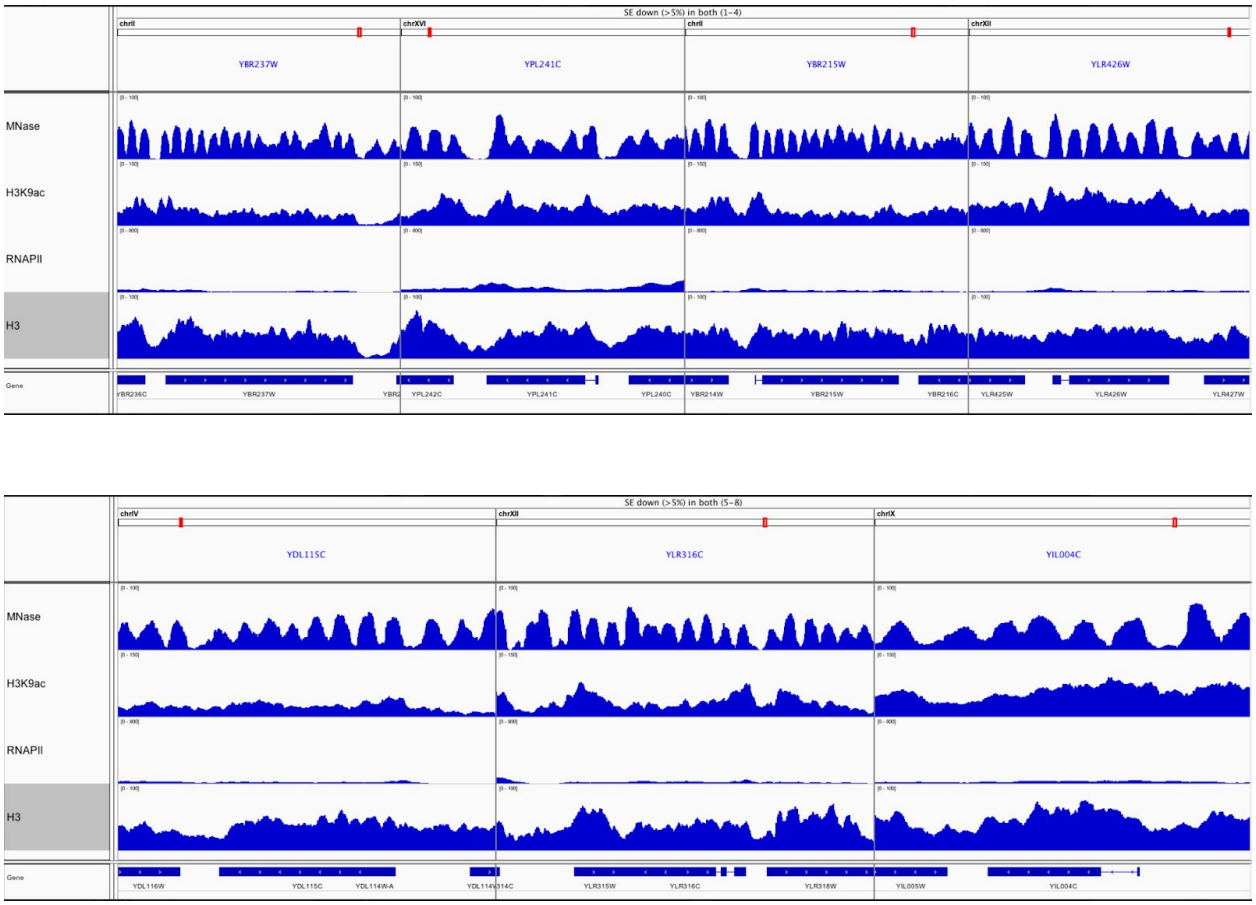


Supplementary Figure 4 (cont'd): Each meta-gene plot represents the average enrichment profile of MNase-seq digestion for individual IC-non-RPGs demonstrating ≥5% decrease in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.

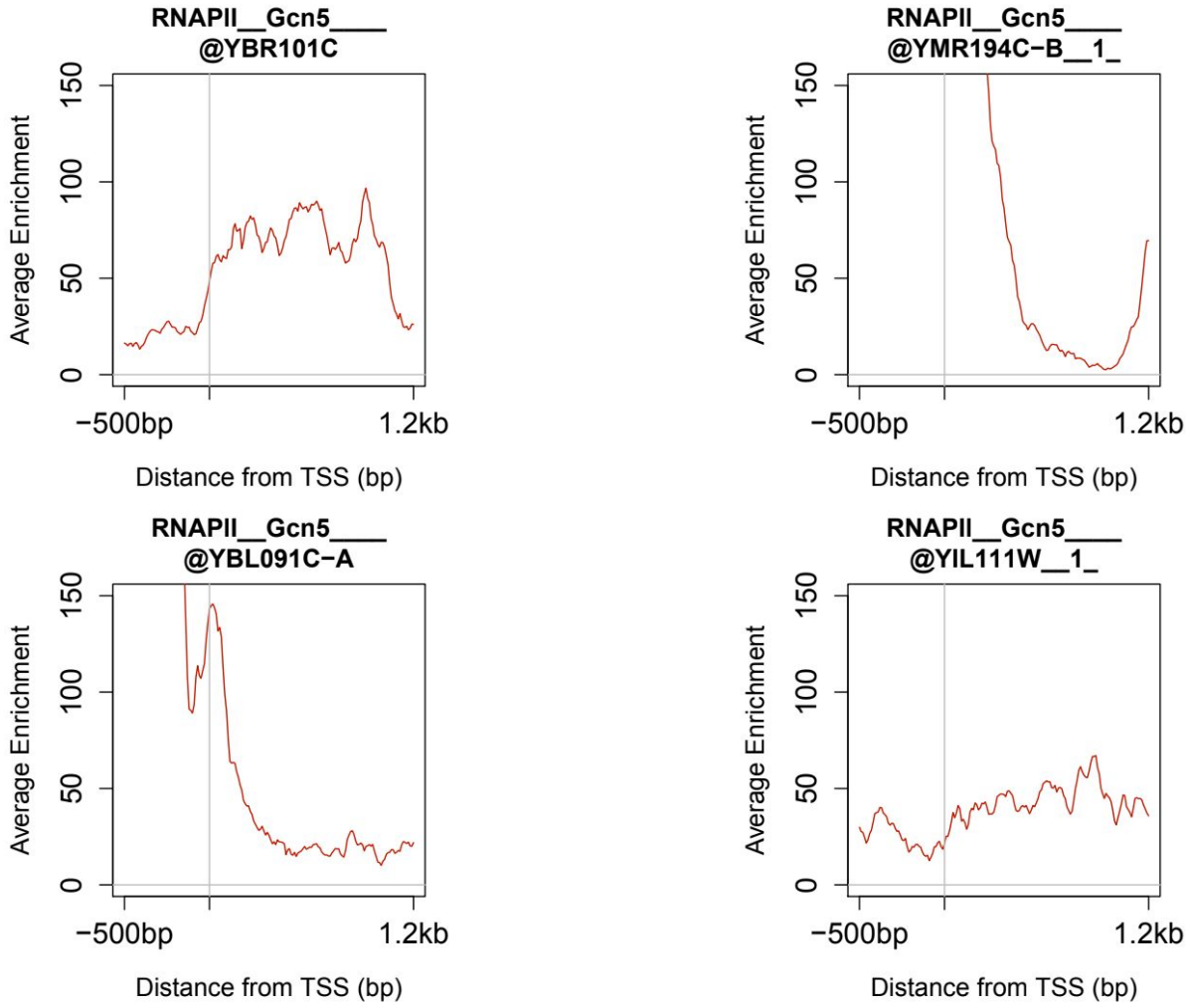




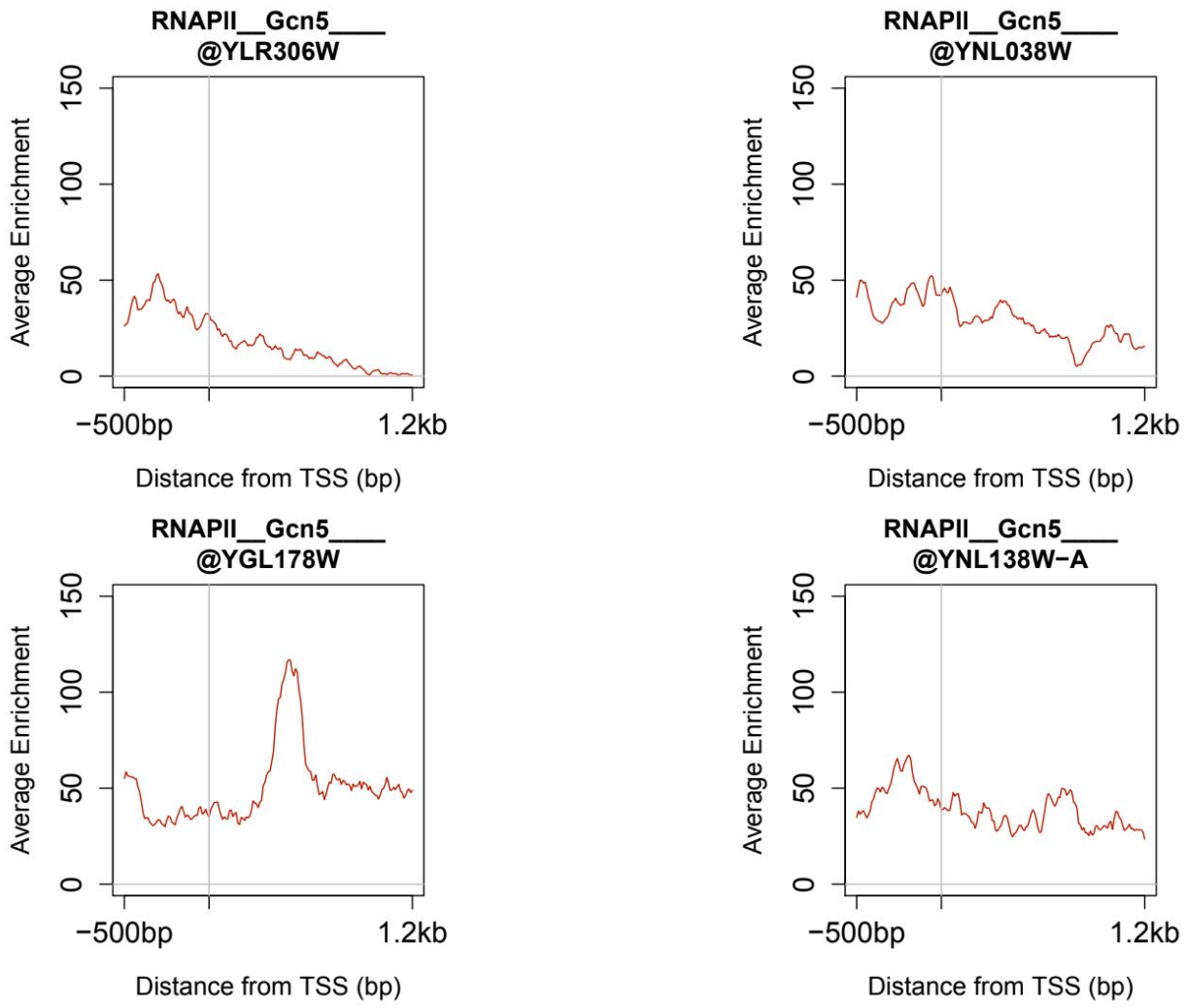
Supplementary Figure 5: Genome browser tracks showing MNase-seq digestion and ChIP-seq profiles of H3K9ac, RNAPII-Ser5p and H3, respectively, for individual IC-non-RPGs demonstrating ≥5% decrease in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while gene locus annotations are represented at the bottom of each track.



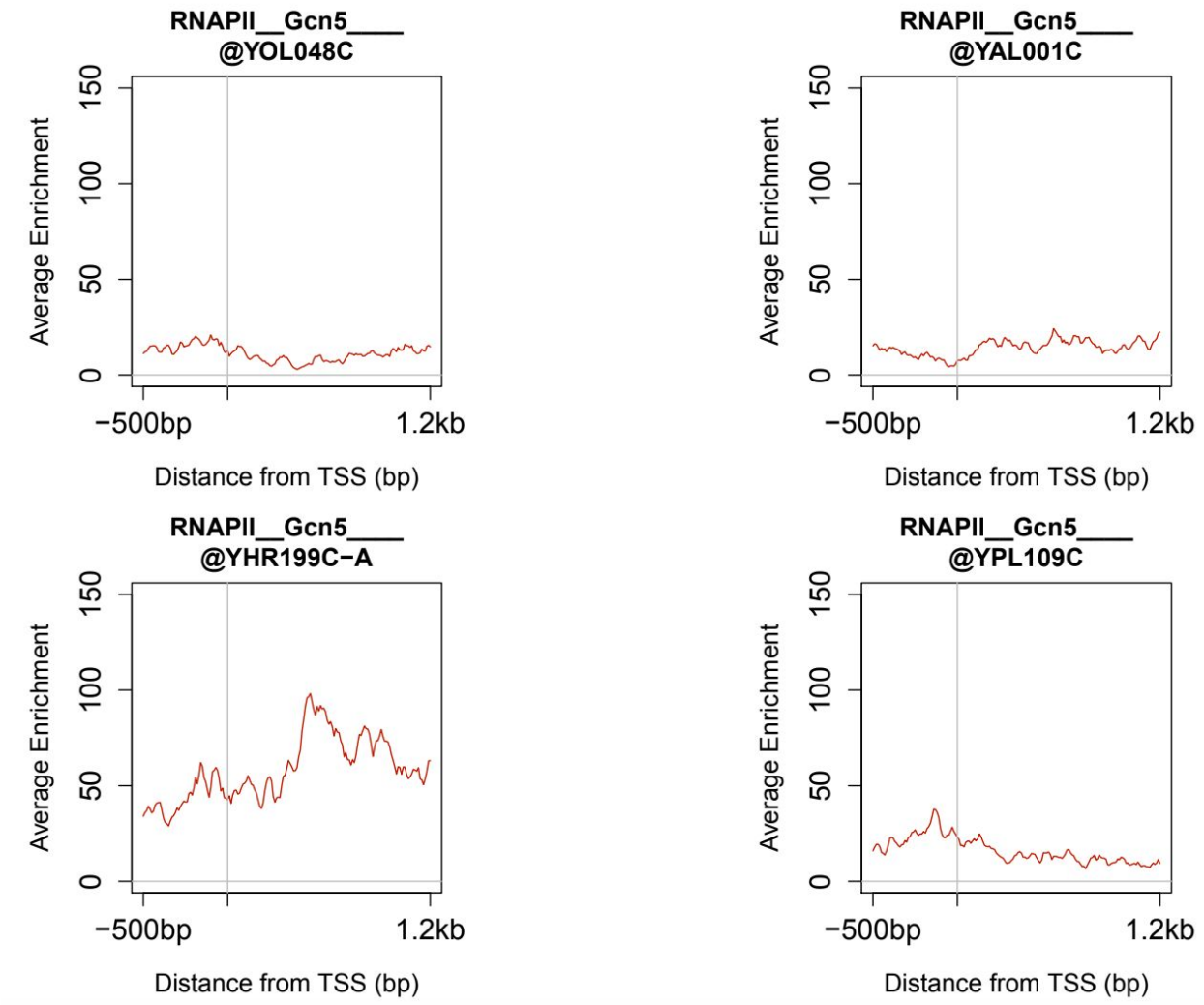
Supplementary Figure 5 (cont'd): Genome browser tracks showing MNase-seq digestion and ChIP-seq profiles of H3K9ac, RNAPII-Ser5p and H3, respectively, for individual IC-non-RPGs demonstrating ≥5% decrease in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while gene locus annotations are represented at the bottom of each track.



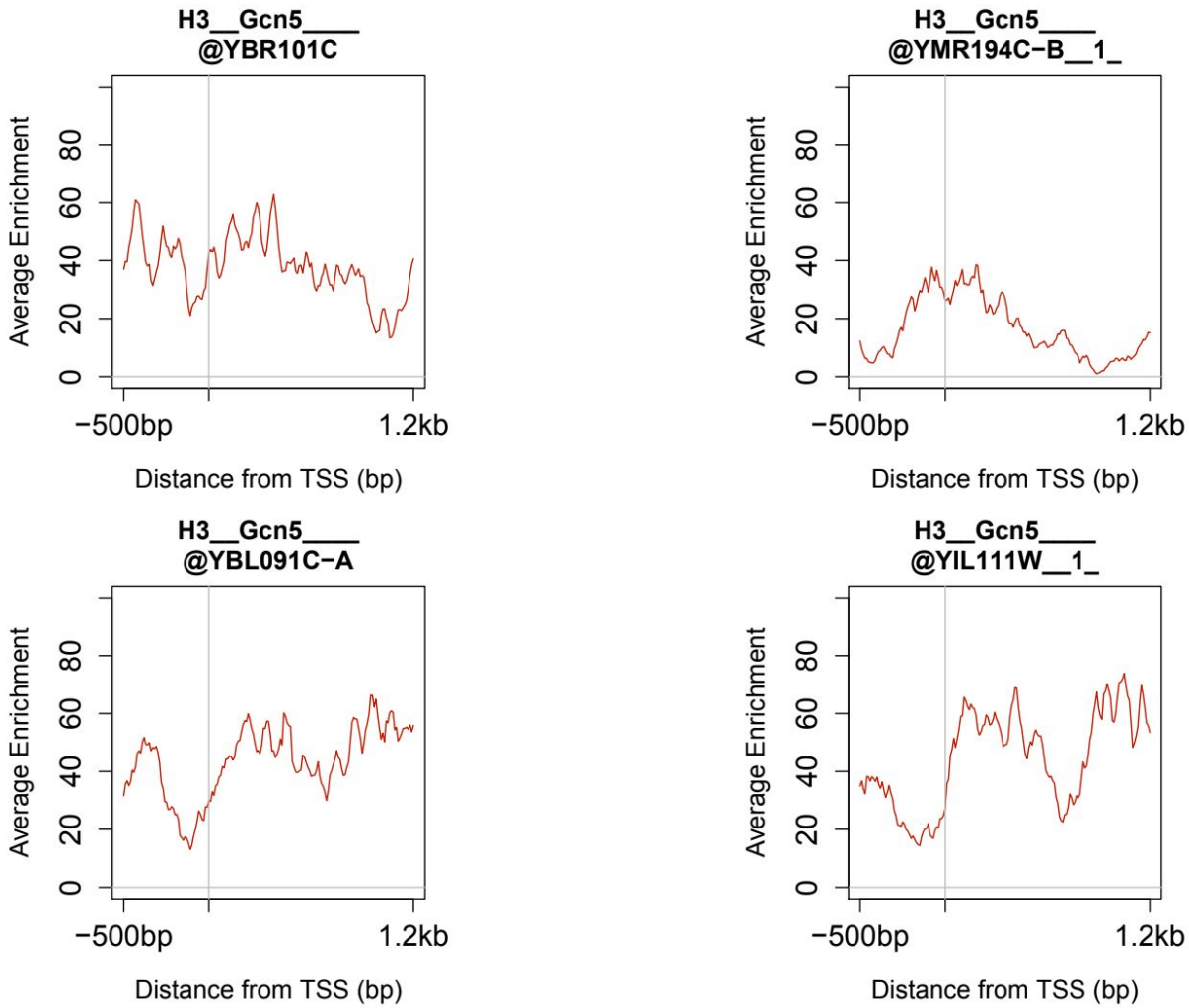
Supplementary Figure 6: Each meta-gene plot represents the average ChIP-seq enrichment profile of RNAPII-Ser5p for individual IC-non-RPGs demonstrating ≥5% increase in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.



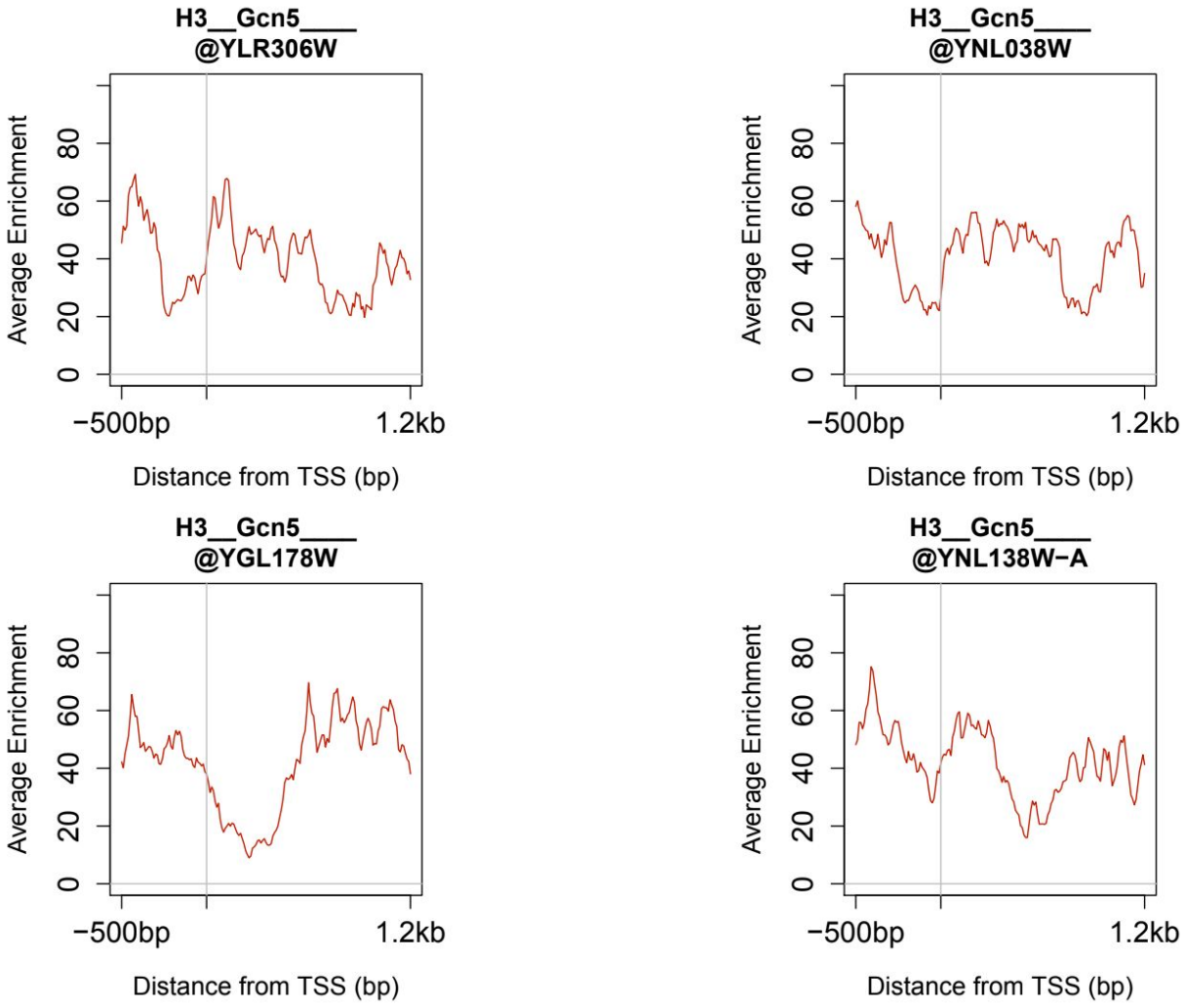
Supplementary Figure 6 (cont'd): Each meta-gene plot represents the average ChIP-seq enrichment profile of RNAPII-Ser5p for individual IC-non-RPGs demonstrating ≥5% increase in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.



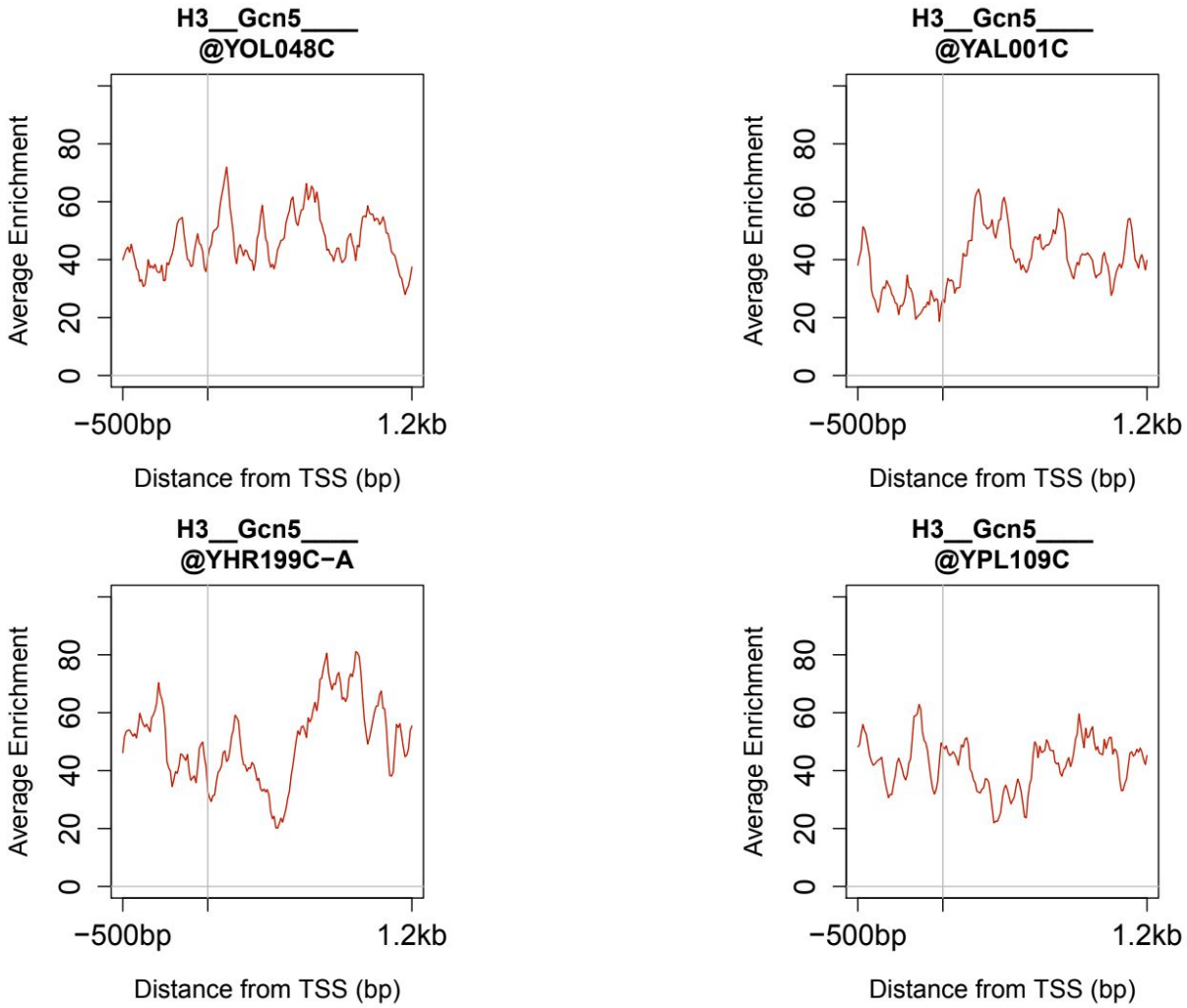
Supplementary Figure 6 (cont'd): Each meta-gene plot represents the average ChIP-seq enrichment profile of RNAPII-Ser5p for individual IC-non-RPGs demonstrating ≥5% increase in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.



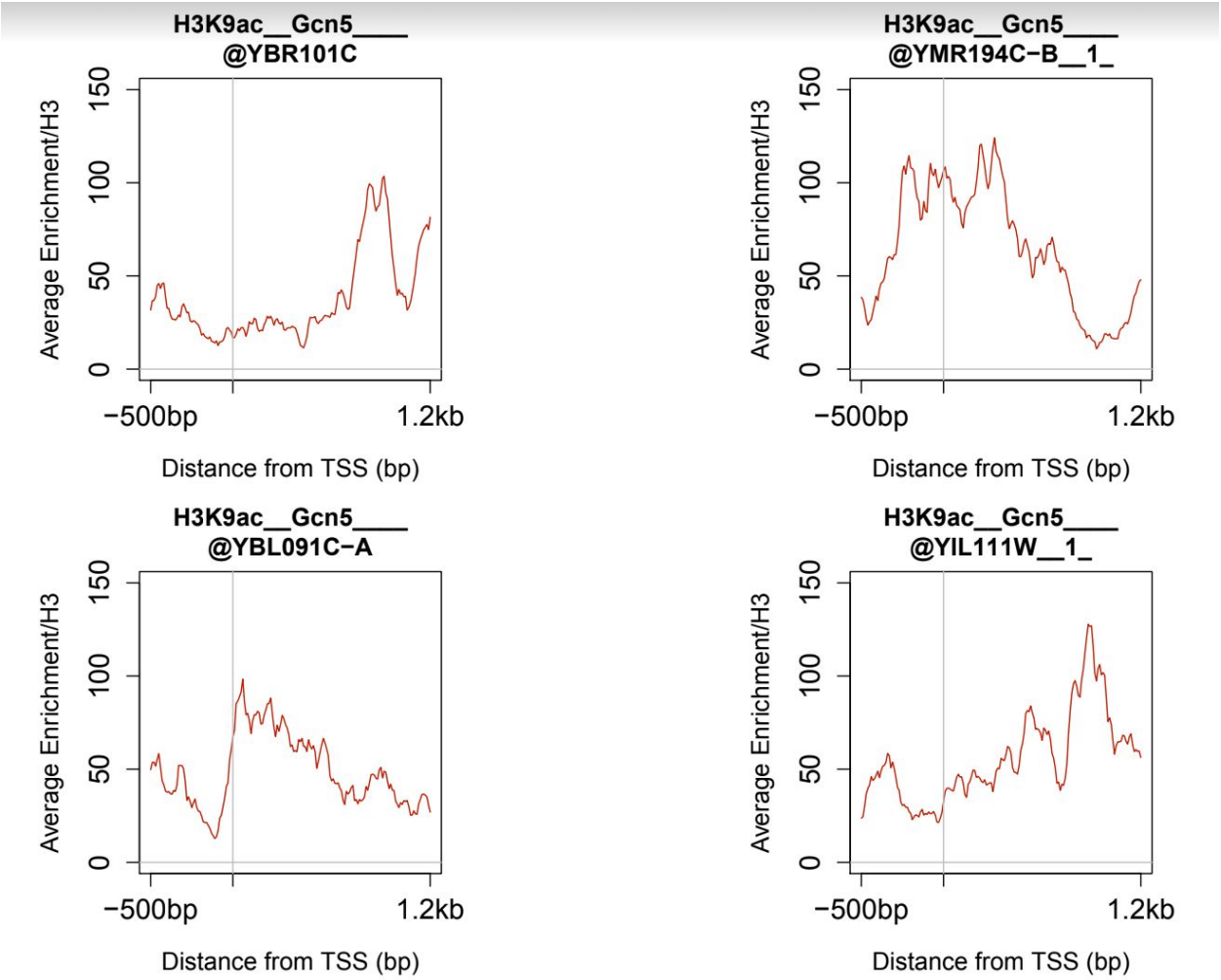
Supplementary Figure 7: Each meta-gene plot represents the average ChIP-seq enrichment profile of H3 for individual IC-non-RPGs demonstrating ≥5% increase in SE under both mutants. Y-axis represents the average enrichment per locus while x-axis represents the distance from the TSS.



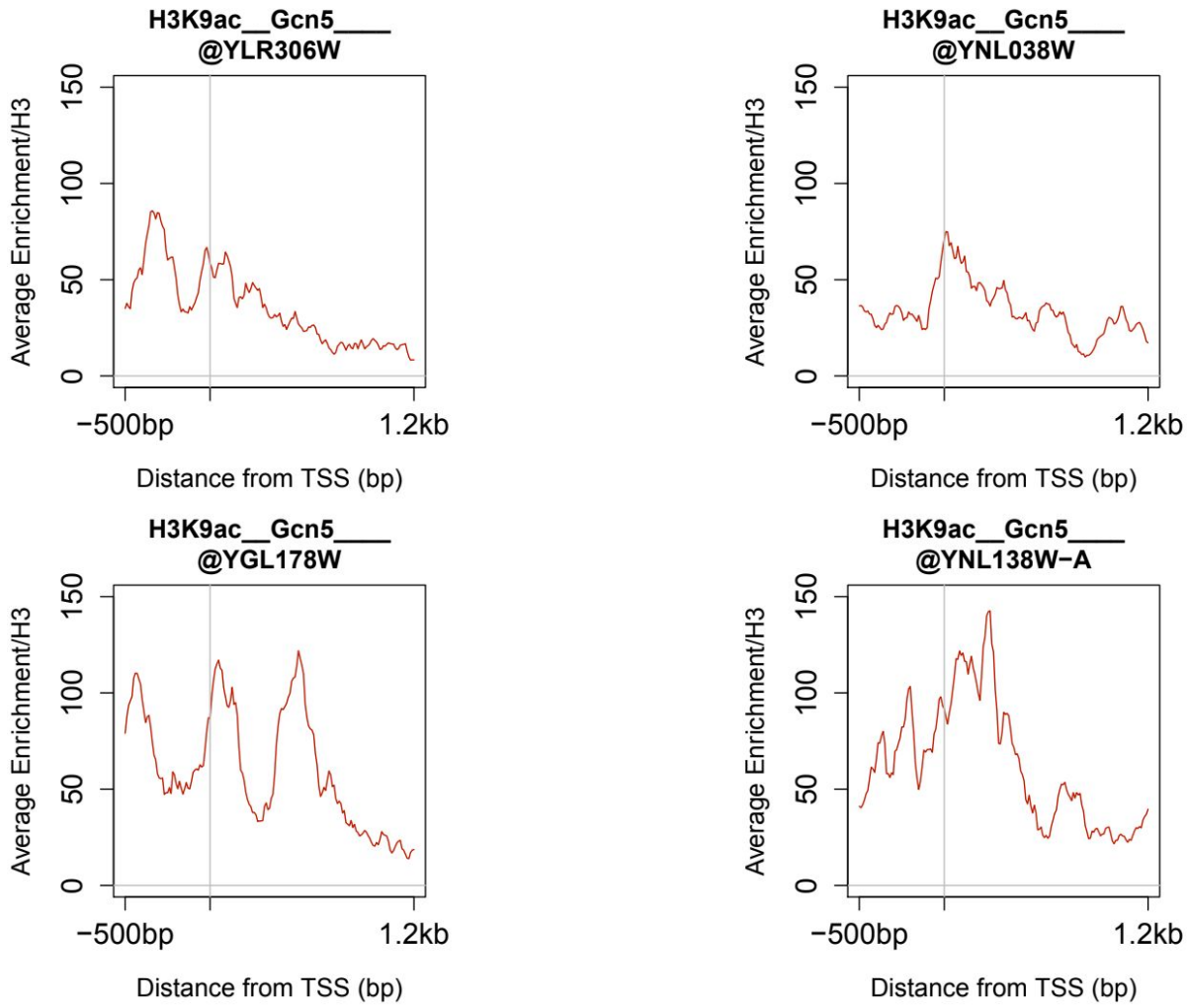
Supplementary Figure 7 (cont'd): Each meta-gene plot represents the average ChIP-seq enrichment profile of H3 for individual IC-non-RPGs demonstrating ≥5% increase in SE under both mutants. Y-axis represents the average enrichment per locus while x-axis represents the distance from the TSS.



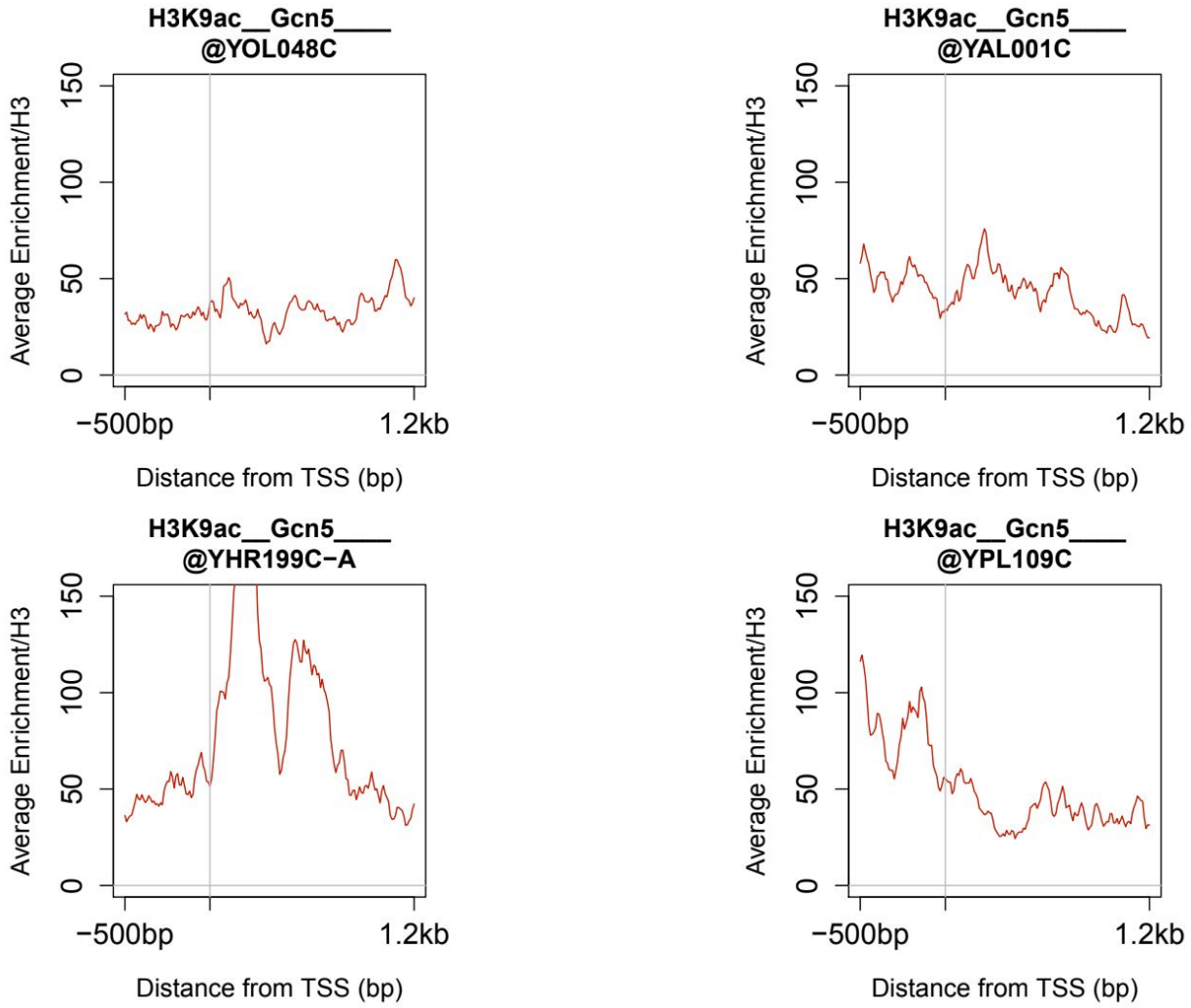
Supplementary Figure 7 (cont'd): Each meta-gene plot represents the average ChIP-seq enrichment profile of H3 for individual IC-non-RPGs demonstrating ≥5% increase in SE under both mutants. Y-axis represents the average enrichment per locus while x-axis represents the distance from the TSS.



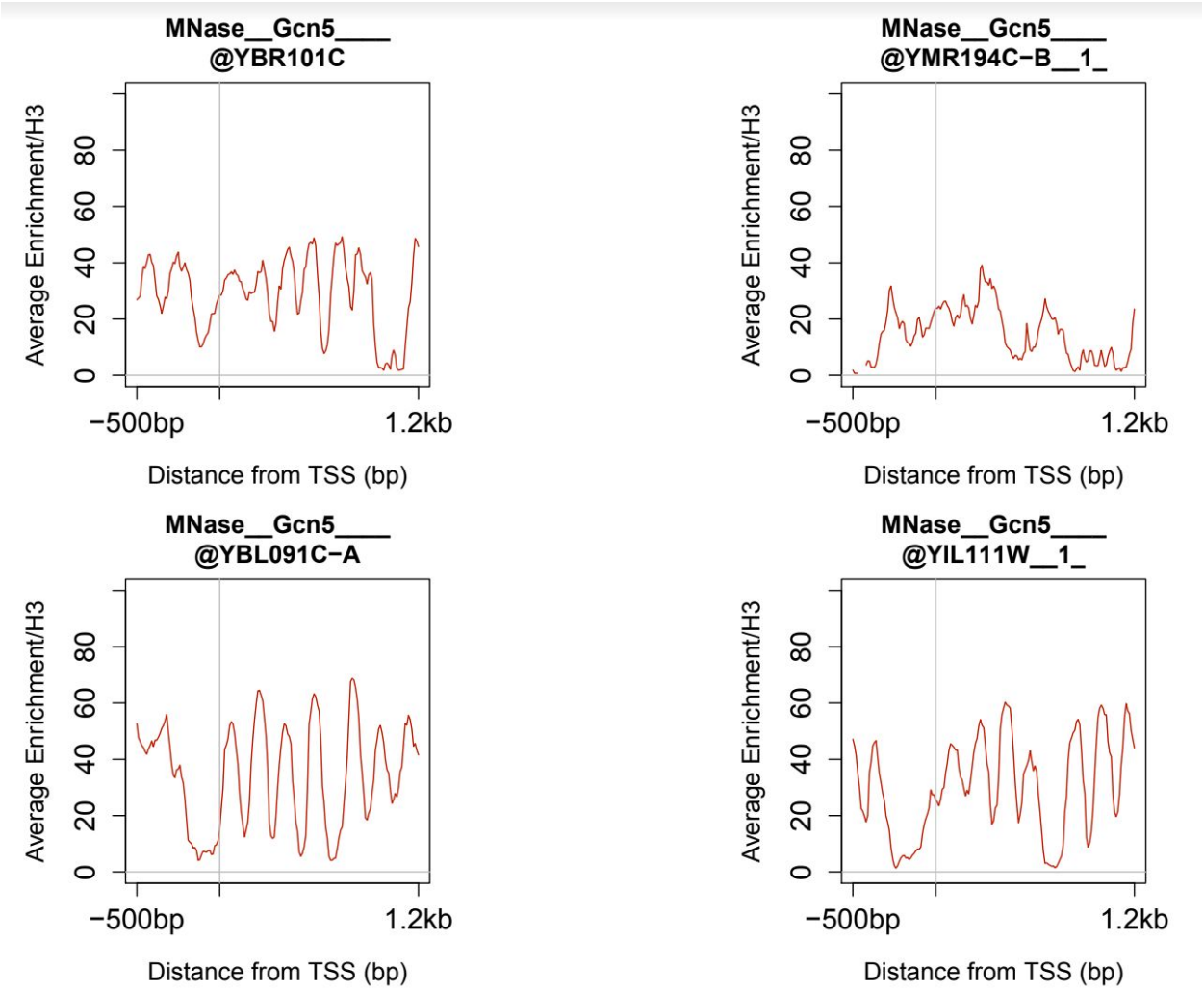
Supplementary Figure 8: Each meta-gene plot represents the average ChIP-seq enrichment profile of H3K9ac for individual IC-non-RPGs demonstrating ≥5% increase in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.



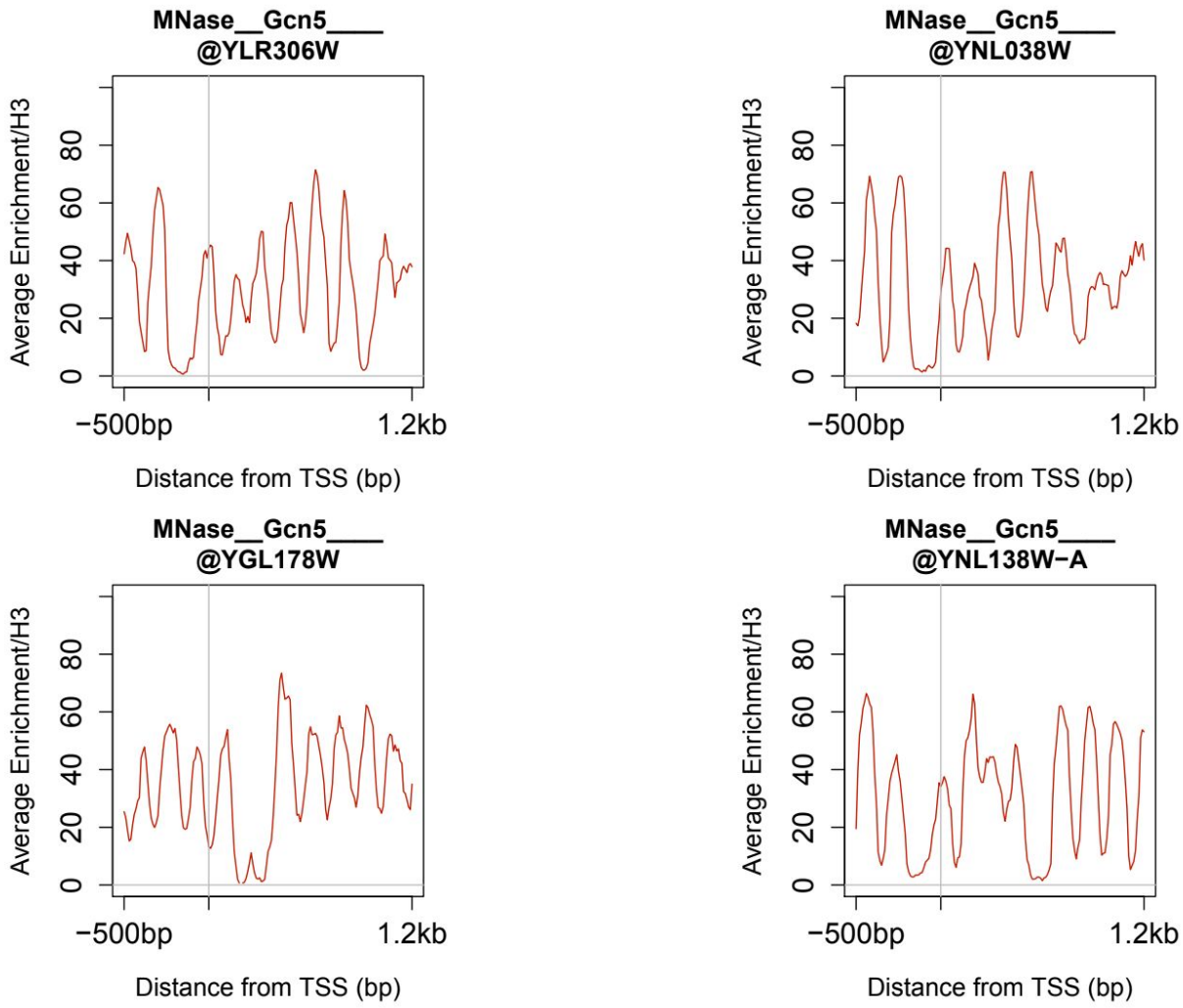
Supplementary Figure 8 (cont'd): Each meta-gene plot represents the average ChIP-seq enrichment profile of H3K9ac for individual IC-non-RPGs demonstrating ≥5% increase in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.



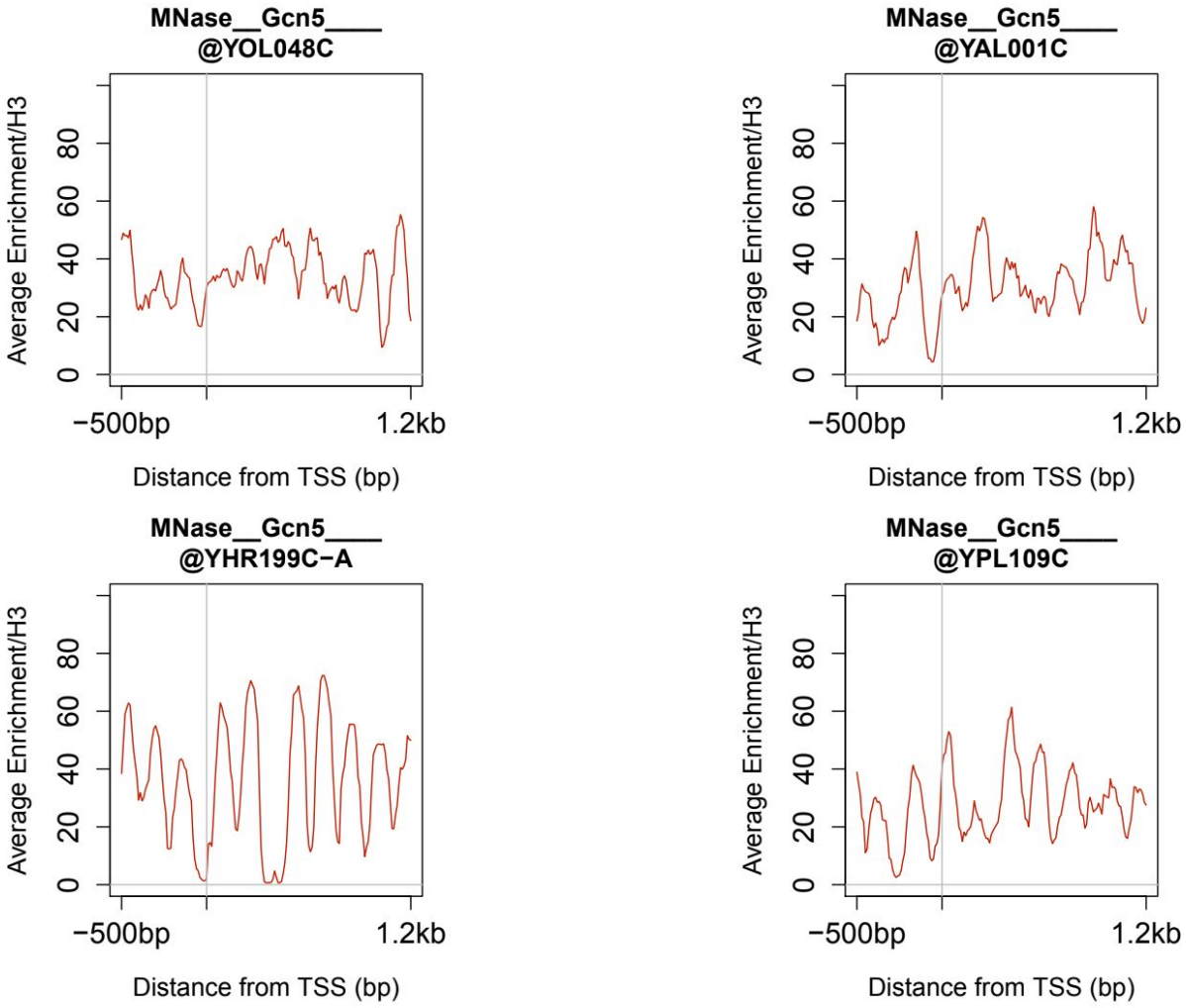
Supplementary Figure 8 (cont'd): Each meta-gene plot represents the average ChIP-seq enrichment profile of H3K9ac for individual IC-non-RPGs demonstrating ≥5% increase in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.



Supplementary Figure 9: Each meta-gene plot represents the average MNase-seq digestion profile of for individual IC-non-RPGs demonstrating ≥5% increase in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.

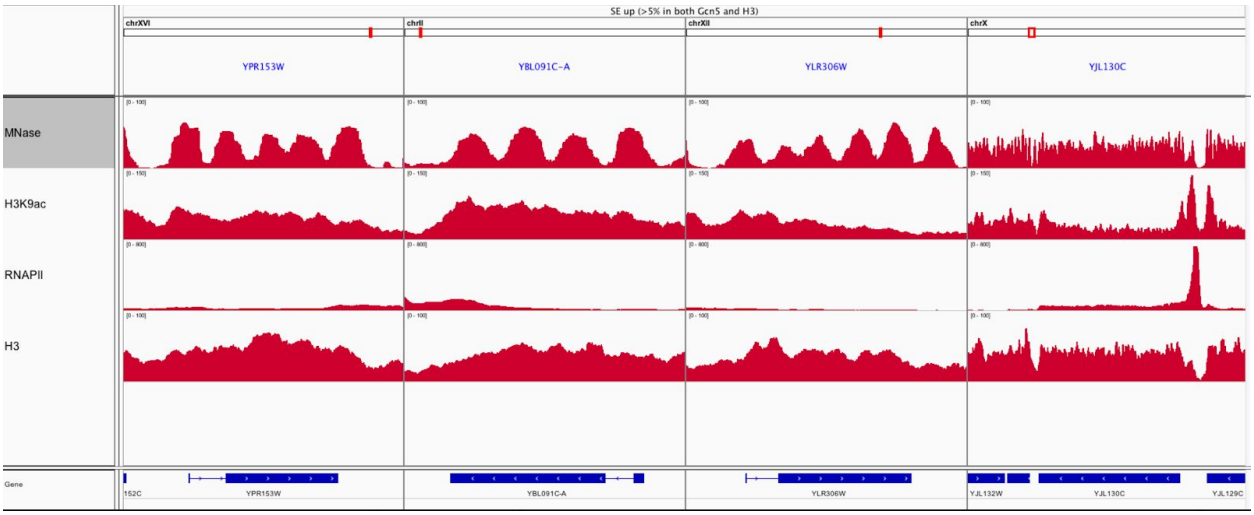


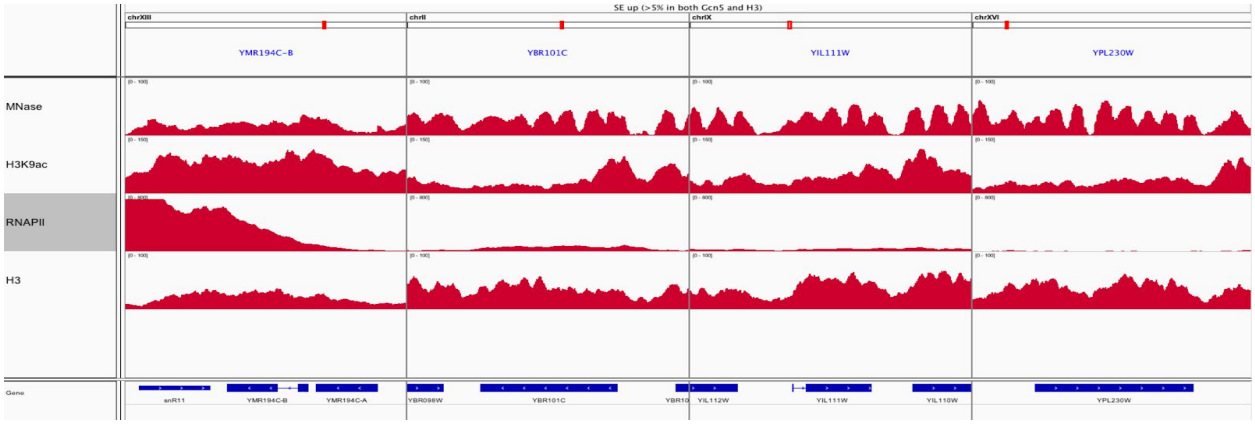
Supplementary Figure 9 (cont'd): Each meta-gene plot represents the average MNase-seq digestion profile of for individual IC-non-RPGs demonstrating ≥5% increase in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.



Supplementary Figure 9 (cont'd): Each meta-gene plot represents the average MNase-seq digestion profile of for individual IC-non-RPGs demonstrating ≥5% increase in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while x-axis represents the distance from the TSS.

Supplementary Figure 10: Genome browser tracks showing MNase-seq digestion and ChIP-seq profiles of H3K9ac, RNAPII-Ser5p and H3, respectively, for individual IC-non-RPGs demonstrating ≥5% increase in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while gene locus annotations are represented at the bottom of each track.





Supplementary Figure 10 (cont'd): Genome browser tracks showing MNase-seq digestion and ChIP-seq profiles of H3K9ac, RNAPII-Ser5p and H3, respectively, for individual IC-non-RPGs demonstrating ≥5% increase in SE under both mutants. Y-axis represents the average enrichment/H3 per locus while gene locus annotations are represented at the bottom of each track.

REFERENCES

Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C., & Komorowski, J. (2009). Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res, 19*(10), 1732-1741. doi:10.1101/gr.092353.109

Awad, A. M., Venkataramanan, S., Nag, A., Galivanche, A. R., Bradley, M. C., Neves, L. T., . . . Johnson, T. L. (2017). Chromatin-remodeling SWI/SNF complex regulates coenzyme Q6 synthesis and a metabolic shift to respiration in yeast. *J Biol Chem, 292*(36), 14851-14866. doi:10.1074/jbc.M117.798397

Boeger, H., Griesenbeck, J., Strattan, J. S., & Kornberg, R. D. (2003). Nucleosomes unfold completely at a transcriptionally active promoter. *Mol Cell, 11*(6), 1587-1598. doi:10.1016/s1097-2765(03)00231-4

Carey, M., Li, B., & Workman, J. L. (2006). RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. *Mol Cell, 24*(3), 481-487. doi:10.1016/j.molcel.2006.09.012

Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., . . . Workman, J. L. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell, 123*(4), 581-592. doi:10.1016/j.cell.2005.10.023

Chandy, M., Gutierrez, J. L., Prochasson, P., & Workman, J. L. (2006). SWI/SNF displaces SAGA-acetylated nucleosomes. *Eukaryot Cell, 5*(10), 1738-1747. doi:10.1128/EC.00165-06

Chatterjee, N., Sinha, D., Lemma-Dechassa, M., Tan, S., Shogren-Knaak, M. A., & Bartholomew, B. (2011). Histone H3 tail acetylation modulates ATP-dependent remodeling through multiple mechanisms. *Nucleic Acids Res, 39*(19), 8378-8391. doi:10.1093/nar/gkr535

Chen, W., Luo, L., & Zhang, L. (2010). The organization of nucleosomes around splice sites. *Nucleic Acids Res, 38*(9), 2788-2798. doi:10.1093/nar/gkq007

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., . . . Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature, 485*(7398), 376-380. doi:10.1038/nature11082

Gopalakrishnan, S., Van Emburgh, B. O., Shan, J., Su, Z., Fields, C. R., Vieweg, J., . . . Robertson, K. D. (2009). A novel DNMT3B splice variant expressed in tumor and pluripotent cells modulates genomic DNA methylation patterns and displays altered DNA binding. *Mol Cancer Res, 7*(10), 1622-1634. doi:10.1158/1541-7786.MCR-09-0018

Henikoff, S. (2016). Mechanisms of Nucleosome Dynamics In Vivo. *Cold Spring Harb Perspect Med, 6*(9). doi:10.1101/cshperspect.a026666

Henriques, R., Magyar, Z., & Bogre, L. (2013). S6K1 and E2FB are in mutually antagonistic regulatory links controlling cell growth and proliferation in Arabidopsis. *Plant Signal Behav, 8*(6), e24367. doi:10.4161/psb.24367

Jimeno-Gonzalez, S., Payan-Bravo, L., Munoz-Cabello, A. M., Guijo, M., Gutierrez, G., Prado, F., & Reyes, J. C. (2015). Defective histone supply causes changes in RNA polymerase II elongation rate and cotranscriptional pre-mRNA splicing. *Proc Natl Acad Sci U S A, 112*(48), 14840-14845. doi:10.1073/pnas.1506760112

Jonkers, I., Kwak, H., & Lis, J. T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife, 3*, e02407. doi:10.7554/eLife.02407

Joshi, A. A., & Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol Cell, 20*(6), 971-978. doi:10.1016/j.molcel.2005.11.021

Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., & Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Res, 32*(Database issue), D493-496. doi:10.1093/nar/gkh103

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X. S., & Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet, 41*(3), 376-381. doi:10.1038/ng.322

Liu, H. L., & Cheng, S. C. (2012). The interaction of Prp2 with a defined region of the intron is required for the first splicing reaction. *Mol Cell Biol, 32*(24), 5056-5066. doi:10.1128/MCB.01109-12

Mason, P. B., & Struhl, K. (2005). Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Mol Cell, 17*(6), 831-840. doi:10.1016/j.molcel.2005.02.017

Matveeva, E. A., Al-Tinawi, Q. M. H., 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. doi:10.1186/s13072-019-0261-1

Neves, L. T., Douglass, S., Spreafico, R., Venkataramanan, S., Kress, T. L., & Johnson, T. L. (2017). The histone variant H2A.Z promotes efficient cotranscriptional splicing in S. cerevisiae. *Genes Dev, 31*(7), 702-717. doi:10.1101/gad.295188.116

Oesterreich, F. C., Herzel, L., Straube, K., Hujer, K., Howard, J., & Neugebauer, K. M. (2016). Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell, 165*(2), 372-381. doi:10.1016/j.cell.2016.02.045

Pavri, R., Lewis, B., Kim, T. K., Dilworth, F. J., Erdjument-Bromage, H., Tempst, P., . . . Reinberg, D. (2005). PARP-1 determines specificity in a retinoid signaling pathway via direct modulation of mediator. *Mol Cell, 18*(1), 83-96. doi:10.1016/j.molcel.2005.02.034

Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., . . . Young, R. A. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell, 122*(4), 517-527. doi:10.1016/j.cell.2005.06.026

Robert, F., Pokholok, D. K., Hannett, N. M., Rinaldi, N. J., Chandy, M., Rolfe, A., . . . Young, R. A. (2004). Global position and recruitment of HATs and HDACs in the yeast genome. *Mol Cell, 16*(2), 199-209. doi:10.1016/j.molcel.2004.09.021

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcarcel, J., & Guigo, R. (2009). Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol, 16*(9), 996-1001. doi:10.1038/nsmb.1658

VanDemark, A. P., Kasten, M. M., Ferris, E., Heroux, A., Hill, C. P., & Cairns, B. R. (2007). Autoregulation of the rsc4 tandem bromodomain by gcn5 acetylation. *Mol Cell, 27*(5), 817-828. doi:10.1016/j.molcel.2007.08.018

CHAPTER 4 - Conclusion and Future Directions

The conversation of how chromatin and pre-mRNA splicing are linked in the cell is widely varied and with the advent of Next Generation Sequencing technologies, it has become an excitingly complex field to explore. The general question I sought to answer with this doctoral thesis is whether or not the major yeast histone acetyltransferase *GCN5* plays a genome-wide role in pre-mRNA splicing. Here, I show for the first time that deletion of *GCN5* or its major histone target region (Lysines 9 through 16 on Histone H3) both resulted in a marked down-regulation of intron containing ribosomal protein gene expression (IC-RPGs). Secondly, I showed for the first time that both *GCN5* and the major histone targets play a role in regulating the efficiency of genome-wide splicing outcomes in yeast. The data supporting these major results are described in Chapter 2.

In both Chapters 2 and 3, I provide support for two mechanisms that may be governing the Gcn5-dependent effects on splicing outcomes genome-wide. Under both mutants studied, splicing outcomes both improved and worsened for intron containing genes. To address the Gcn5-dependent increases in splicing outcomes, I showed that *GCN5* deletion decreases RPKMs of all IC-RPGs, which triggers a redistribution of the limited spliceosomes and thus increased splicing to IC-non-RPGs. This mechanism was first elucidated in a non-Gcn5 specific context with Munding et al., 2013, and confirmed for the first time for Gcn5 in my RNA-seq data-set. To address the ICGs that were decreasing in splicing efficiency and effectively escaping the aforementioned “RPG effect”, I used publicly available ChIP-seq and MNase-seq data from Bruzzone et al., 2018 to survey the chromatin landscape of these genes (described in Chapter 3). To date, this was the first time MNase-seq data was used to answer questions about the

chromatin environment and its connection to splicing in a Gcn5 context. I demonstrated that ICGs that are either increasing or decreasing in splicing efficiency in a Gcn5-dependent context are markedly differentiated by distinct chromatin signatures. In combination, I have opened an array of possibilities for mechanisms regarding the genome-wide role of Gcn5 in splicing.

Some of the major strengths of this dissertation research lie squarely the methodological approach to uncover novel insights about chromatin and splicing in a Gcn5-dependent context. To date, RNA-seq combined with splicing-relevant data analysis methods (i.e. splicing efficiency calculations) have not been used to understand the histone acetyltransferase role of Gcn5 in splicing. Additionally, using publically available data is a powerful and efficient approach to answering complex questions about gene regulation. As the questions regarding the relationship between chromatin and splicing become more sophisticated, the collaborative approach of asking new and different questions with the wealth of data that already exists in the scientific community can efficiently reduce the cost-benefit ratio of NGS-driven projects. For example, the data provided by Bruzzone et al., 2018 was not originally generated for the purposes of understanding the relationship between chromatin and splicing. However, I have been able to creatively "read-in-between" the lines in my initial observations from Chapter 2, generate hypotheses about the categories of genes I am interested in (SE up and SE down in Gcn5-dependent context), while imposing the publicly available data as a cost-effective and powerful tool to test my hypotheses. With this new approach, I have been able to show that there is a particularly unique chromatin environment distinguishing non-RPGs that are increasing and non-RPGs are decreasing in Gcn5-dependent splicing efficiency.

The main limitations of this research lie in supplementing the bioinformatics data with classic biochemistry experiments as a way to dive deeper into the mechanism. Performing Co-Immunoprecipitation experiments in a Gcn5 +/- context between H3K9ac and U2 snRNP factors would provide stronger evidence for a direct physical relationship between the chromatin template and spliceosome assembly. Additionally, observation of genetic interactions between Gcn5, the major histone targets and additional spliceosome components can be of support to this mechanism. Another limitation that can be addressed with a future direction is to perform NET-seq experiments to explore the role of RNAPII pausing in this context. Given the conclusions regarding differences in the chromatin environment per Gcn5-dependent splicing outcomes, understanding the nature of RNAPII in a Gcn5 +/- context at these gene loci will help to refine the mechanism at play. Additionally, using slow or fast RNAPII mutants in combination with *GCN5* deletion mutants as a background for the metagene analysis can add more nuance to the whether a kinetic mechanism may be a result of the chromatin signature differences in these groups of genes. Another area of future opportunity in this project would be to add additional Gcn5 and H3 mutants to the array of strains tested. Since Gcn5p interacts with other SAGA-complex units, it would be interesting to repeat the above experiments in SAGA-complex mutants as well as H4 acetyltransferases to differentiate HAT-specific Gcn5 from the influence of SAGA or histone acetyltransferases in general. Additionally, since Gcn5 has multiple targets outside of the region focused on in this dissertation, it would be of interest to examine point mutants of H3 lysines within the 9-16 region as well as outside of the region. This would also provide more information about the influence of specific residues to facilitating the relationships described in the previous chapters.